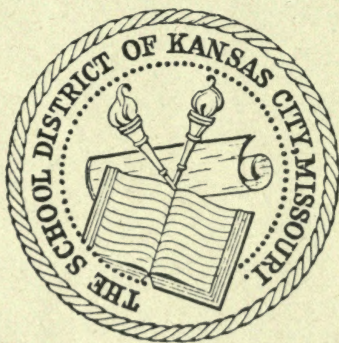

Kansas City
Public Library



This Volume is for
REFERENCE USE ONLY

PUBLIC LIBRARY
KANSAS CITY
MO

YACOBUS 3.15.19
VTD 24.12.19
OM

PUBLIC LIBRARY
KANSAS CITY
MO

WAGG LEE
THE COUNTRY
OF

V. 32! Jan - Nov, 1953

H E B E L L S Y S T E M

Technical Journal

VOTED TO THE SCIENTIFIC AND ENGINEERING
PECTS OF ELECTRICAL COMMUNICATION

LUME XXXII

JANUARY 1953

NUMBER 1

FEB 3 - 1953

Surface Properties of Germanium

WALTER H. BRATTAIN AND JOHN BARDEEN

1

Frequency Economy in Mobile Radio Bands

KENNETH BULLINGTON 42

Intermodulation Interference in Radio Systems

WALLACE C. BABCOCK 63

Magnetic Resonance: Part I—Nuclear Magnetic Resonance

KARL K. DARROW 74

Delay Curves for Calls Served at Random

JOHN RIORDAN 100

The Evaluation of Wood Preservatives — Part I

REGINALD H. COLLEY 120

Motion of Gaseous Ions in Strong Electric Fields

GREGORY H. WANNIER 170

Abstracts of Bell System Papers Not Published in this Journal 255

Contributors to this Issue 261

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

S. BRACKEN, *President, Western Electric Company*

F. R. KAPPEL, *Vice President, American Telephone
and Telegraph Company*

M. J. KELLY, *President, Bell Telephone Laboratories*

EDITORIAL COMMITTEE

E. I. GREEN, *Chairman*

A. J. BUSCH

F. R. LACK

W. H. DOHERTY

J. W. MCRAE

G. D. EDWARDS

W. H. NUNN

J. B. FISK

H. I. ROMNES

R. K. HONAMAN

H. V. SCHMIDT

EDITORIAL STAFF

PHILIP C. JONES, *Editor*

M. E. STRIEBY, *Managing Editor*

R. L. SHEPHERD, *Production Editor*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; Alexander L. Stott, Treasurer. Subscriptions are accepted at \$3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXII

JANUARY

NUMBER 1

Copyright, 1953, American Telephone and Telegraph Company

Surface Properties of Germanium

By WALTER H. BRATTAIN* and JOHN BARDEEN†

(Manuscript received September 3, 1952)

The contact potential (c.p.) and the change of contact potential with illumination $(\Delta c.p.)_L$ of several germanium surfaces have been measured. The reference electrode used was platinum. It was found that the c.p. could be cycled between two extremes about 0.5 volts apart by changing the gaseous ambient. Ozone or peroxide vapors gave the c.p. extreme corresponding to the largest dipole at the Ge surface. Vapors with OH radicals produced the other extreme. There is a one to one correlation between c.p. and $(\Delta c.p.)_L$. For 12-ohm cm n-type Ge $(\Delta c.p.)_L$ was large and positive when the surface dipole was largest, decreased to zero and became slightly negative as the surface dipole decreased to its smallest value. The variation for 12-ohm cm p-type Ge was just opposite as regards both sign and dependence on surface dipole. The surface recombination velocity was found to be independent of c.p. For a chemically prepared surface it was 50–70 cm/sec and 180–200 cm/sec for n and p-type surfaces respectively. A theory is given that explains the results in terms of surface traps, N_a per cm^2 donor-type traps near the conduction band and N_b per cm^2 acceptor-type traps near the filled band. A quantitative fit with experiment is obtained with only one free parameter. The results are direct evidence for the existence of a space charge layer at the free surface of a semiconductor.

INTRODUCTION

Every one is familiar with the fact that it is necessary to expend energy to remove an electron from a conducting solid. This energy is

* Bell Telephone Laboratories.

† University of Illinois. The contributions of the second author to this work started while he was a Member of the Technical Staff of Bell Telephone Laboratories and continued at the University of Illinois.

called the work function. The work function is caused in part by a charged double layer or dipole at the solid surface. In metals this dipole extends over a distance of the order of 10^{-8} cm. In semiconductors however part of the dipole extends into the semiconductor to a distance of the order of 10^{-6} to 10^{-4} cm depending on the properties of the semiconductor. This part of the surface dipole is called the space-charge layer. The rest of the surface dipole has approximately the same extent as in metals.

In the space between any two conducting solids there is a contact potential caused by the difference between the work functions of the two surfaces. One can measure this contact potential by several methods. We have used an adaptation of the well known method of Kelvin. If one has a reference electrode whose work function remains constant, then by measuring the c.p. between this electrode and another surface one can measure any changes in work function or total dipole of this second surface.

The above method has been used to study the properties of the germanium surface in a gaseous ambient at atmospheric pressure. It has been found that the total dipole at the germanium surface can be changed by changing the ambient and further that by proper control of the ambient the surface can be cycled back and forth between two extremes of small or large dipole corresponding to a c.p. change or work function difference of the order of one-half volt.

If one upsets the thermal equilibrium in the germanium by creating excess electron-hole pairs near the surface, the potential of the surface will change until a steady state is reached. When the extra electron-hole pairs are introduced by illuminating the surface with light, the potential change shows up as a measurable change in contact potential between the reference electrode and the Ge surface.¹ It has been found that this contact potential change on illumination $(\Delta c.p.)_L$ is large and positive on *n*-type Ge when the surface dipole is large, then decreases to zero and becomes slightly negative as the surface dipole decreases to the smaller extreme. For *p*-type Ge the $(\Delta c.p.)_L$ is large and negative when the surface dipole is small and goes through zero and becomes slightly positive as the surface dipole increases to the larger extreme.

One can describe qualitatively what is going on as follows. The extra hole and electron pairs created by the light diffuse either to the interior or to the surface to recombine. The recombination in the interior is governed by the body life time τ . The surface recombination is characterized by a recombination velocity v_s . When the surface is illuminated its potential, with respect to the interior, changes until the combined

flow of holes and electrons to the surface and interior is just equal to the rate they are being created by the light. The sign and magnitude of the potential change for a given illumination depends on the body properties of the germanium and on the size of the space charge layer.

These experimental results are direct evidence for the existence of a space charge layer at the free surface of a semiconductor. They not only confirm the results obtained for silicon surfaces¹ but go much further in that they enable one to determine how the layer is changed by the gaseous ambient used.

It is known that the surface recombination velocity, v_s , can be changed, by large factors, by surface treatment.² For mechanically treated surfaces v_s approaches thermal velocities. Every hole or electron striking the surface recombines. For such a surface it is found that $(\Delta c.p.)_L$ is too small to be measured. On the other hand v_s can be as low as 100 cm/sec for chemically polished or etched surfaces such as those used in the experiments where $(\Delta c.p.)_L$ was measured. In this case one wishes to know how v_s depends on the gaseous ambient. This was measured for the same surface used in measuring $(\Delta c.p.)_L$ and it was found that, for the ambients used, v_s is approximately a constant and therefore independent of the other surface changes.

A quantitative theory, some details of which are in the Appendix, has been formulated to explain the results. It is proposed that there are two types of recombination traps at the surface: donor type, N_a per cm², with energies, E_a , near the conduction band and acceptor type, N_b per cm², with energies, E_b , near the filled band. Surface recombination takes place by electrons and holes successively going into one of the two types of traps. To account for the fact that v_s is unchanged by changes in ambient, it is assumed that the concentrations of these traps are independent of ambient. Changes in c.p. with ambient are assumed to result from adsorption and desorption of fixed ions which are at an effective distance $\ell \sim 2 \times 10^{-7}$ cm outward from the surface traps. A schematic energy level diagram is given in Fig. 13, to be discussed later.

The charge of the ions is compensated mainly by charges in the surface traps which, together with the ions, form a double layer. A large part of the change in c.p. with ambient results from changes in this double layer. There is also a change in barrier height, $-eV_B$, associated with the redistribution of electrons in the traps. An increase in negative ions on the surface requires a decrease in number of electrons in traps, and thus a higher barrier.

Part of the change with light, $(\Delta c.p.)_L$, occurs in the body of the semiconductor and part occurs across the barrier layer. Changes in V_B

and occupancy of the traps compensate in their effect on surface recombination, so that v_s is unchanged by ambient. This means that changes in concentrations of electrons and holes in the interior with illumination and thus the body contribution to $(\Delta c.p.)_L$ are independent of ambient. Changes in $(\Delta c.p.)_L$ with ambient result solely from changes in V_B and are in the directions we have described earlier.

The concentrations of carriers, n and p , and their change with light are obtained as follows. Drift mobilities, μ_n and μ_p , and the equilibrium product, np , are known from earlier experiments of G. L. Pearson, J. R. Haynes and W. Shockley.³ From these, and the resistivity of the sample, which was measured, n and p can be determined. The light source is calibrated in terms of hole-electron pairs created per cm^2 per sec. The recombination rate is determined from the body life time τ and the surface recombination velocity. From these latter three measurements, one can calculate the steady state density of electrons and holes near the surface when it is illuminated.

Mention should be made here of the fact that oxygen has been found to play a definite role on the Ge surface. The large extreme in dipole is obtained when active oxygen (ozone) is introduced into the gaseous ambient. Peroxide vapors have the same effect. The other extreme is produced by vapors having an OH radical, water vapor, alcohol etc. A number of vapors not falling in either of the above classes have little or no effect on the surface dipole. Another result is that the difference in work function or dipole between n - and p -type Ge is small. This is to be expected from previous work.

We shall first discuss the experimental technique and then give the experimental results. The main conclusions of the theory will then be outlined and compared with these results.

EXPERIMENTAL METHOD

The Ge surface to be measured and the reference electrode are mounted under a bell jar. Oxygen or nitrogen as desired is allowed to flow through the bell jar at a rate of approximately 2 liters per min. The volume of the bell jar is 16 liters. The gas used flows over a drying column of silica-gel and then calcium chloride. Means are provided for bubbling this gas through any desired liquid before it enters the bell jar. A spark discharge can be run in the gas flow line. The reference electrode is placed parallel to the Ge surface about 1 mm away. It is mounted on a vibrating reed which is driven, electromagnetically, at its resonant frequency of about 90 cycles per second and at an amplitude of the order of 0.1 mm. This varies the capacity sinusoidally giving rise to an electrical signal when any potential,

contact or other, exists between the surfaces. When the proper dc potential is applied between the surfaces this signal goes to zero. If no other potentials are present this dc potential is equal and opposite to the contact potential. A phase reference method* is employed to determine this balance point with a relative accuracy of $\pm 5 \times 10^{-4}$ volts. A diagram of the Ge-reference electrode circuit is shown in Fig. 1. Care must be taken to shield this circuit. Stray capacity reduces sensitivity and should be minimized. Charged insulators inside the shield will produce an apparent c.p. All conducting surfaces other than the Ge should be relatively far from the moving reference electrode. The surface is illuminated through a compound lens system by focusing the filament image, of a suitable projection bulb, on the germanium surface. This light passes through the grid of the reference electrode which removes about 10 per cent of the light. The light can be modulated by a square wave chopper, so that $(\Delta c.p.)_L$ can be measured on an ac basis. Both Ta and Pt reference electrodes have been used. The Pt electrode appears to be somewhat more constant. If the Ge surface is replaced by a gold electrode the contact potential difference is practically independent of the changes in the gas ambient. The arrangement is such that two samples can be mounted in the bell jar and the reference electrode moved from one to the other without opening the bell jar. In this way two surfaces can be compared, without any question arising of long time drifts in the reference electrode.

EXPERIMENTAL RESULTS

I. Change of c.p. between Ge and Pt reference electrodes as a function of the gaseous ambient

When this work was started the object was to find some means of varying the c.p. It was thought that the actual values of the c.p. would

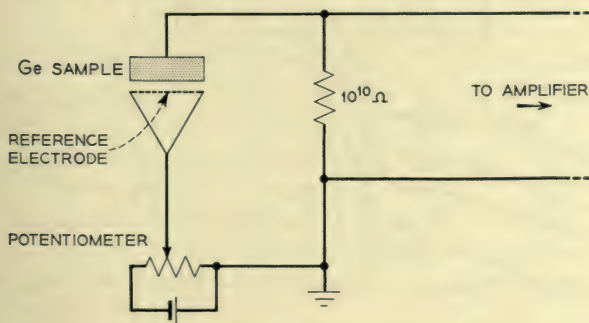


Fig. 1 — Schematic of experimental circuit.

* H. R. Moore designed and made the electronic equipment used to do this.

be highly dependent on past history of the Ge sample. If however one could measure one or more other properties, such as $(\Delta c.p.)_L$ and v_s , on the same surface at the same time, then one could look for correlations between these properties. In this manner one might be able to eliminate the past history as a factor. From previous experiments in the highly variable ambient of room air we knew the order of magnitude of the variation to be expected. At first it was impossible to produce this range of variation under the bell jar. The contact potential always drifted in the direction of a positive extreme, i.e., small total dipole at the Ge surface. The only way found to get a really large change in the opposite direction was to lift the bell jar and expose the sample to room air. These phenomena were finally traced to the presence of negative ions, possible salt ions, in the room air. These were not present in the oxygen and nitrogen supplies we used for creating the ambient in the bell jar.

It was found that the opposite c.p. extreme could be produced, under the bell jar, by running a spark discharge in dry oxygen as it was flowing into the system. The next step was to cycle the c.p. from one extreme to the other and back again. The procedure was to start with the spark discharge in dry oxygen, change to either wet O_2 or wet N_2 and to end with dry O_2 . The development of this dependable and reproducible cycle was a great aid to the proposed study. Fig. 2 is a plot of contact potential versus time for a single crystal slice D of Ge cut from a melt that was p -type. The surface was prepared by removing some of the Ge with a silicon carbide (180 mesh) blast of approximately ten pounds air pressure. The Ge was then mounted in the bell jar within one-half minute of the "sandblast," and the dry O_2 flow started. The c.p. was followed for a few minutes to be sure everything was working properly.

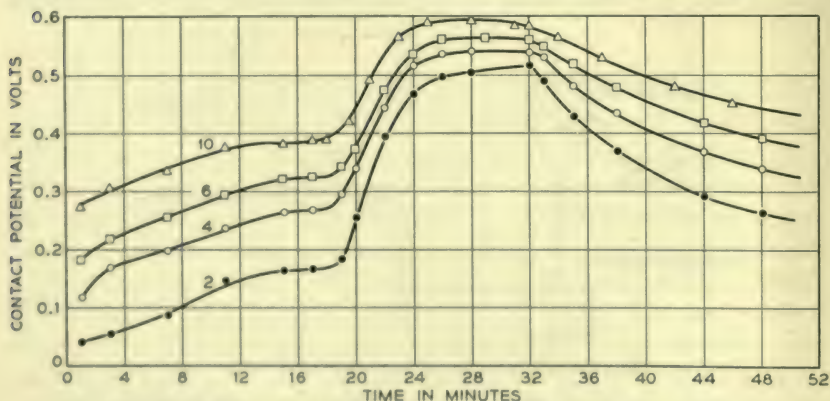


Fig. 2 — Contact potential cycles for sandblasted sample D .

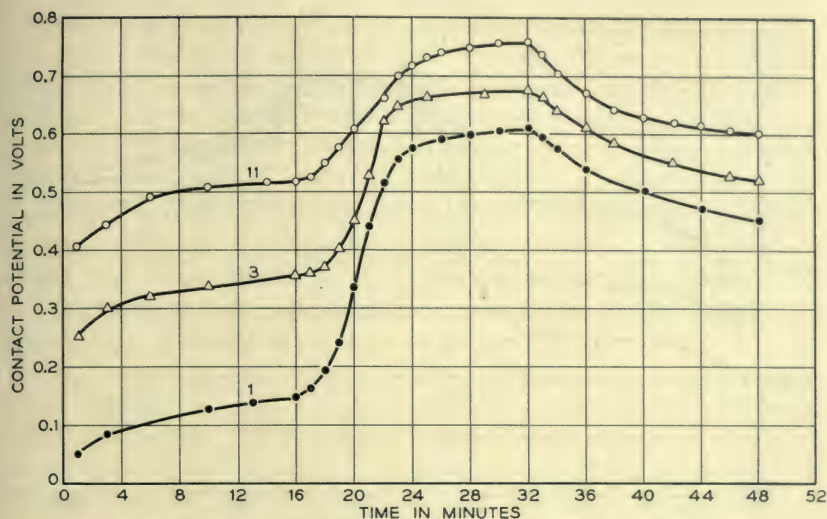


Fig. 3 — Contact potential cycles for etched sample A.

In Fig. 2 zero time is taken after the spark discharge was run in the O_2 flow line for 2 minutes. This started the first cycle. After approximately 17 minutes the O_2 was made to bubble through H_2O . Fifteen minutes or so later the O_2 was changed back to dry. At this time, 32 minutes, the flow rate was increased by a factor of three. The c.p. was followed for about 17 minutes, then the process was repeated. The results and the reason for the choice of time intervals, are all evident from a study of Fig. 2. The spark discharge in O_2 decreases the c.p. After this treatment the c.p. increases with time, most of the change occurring in the first 15 minutes. The wet O_2 then increases the c.p. to a maximum value which is reached in about 15 minutes. Finally the dry O_2 reduces the c.p. It is evident that there is a quasi-equilibrium value of c.p. in dry O_2 , to which the c.p. returns after either extreme treatment. At first there is quite a large shift from cycle to cycle but this shift gradually disappears as the cycling is continued. In Fig. 2 cycles 2, 4, 6 and 10 are shown. Very little change takes place after cycle 10. Such results have been obtained many times over a period of two years. When allowance is made for shifts in work function of the reference electrode and for the fact that the experimental technique improved as the work progressed it is found that all the results for a given sample of Ge are very consistent.

In Fig. 3 are shown similar results for an *n*-type slice A. In this case the surface was first ground or sandblasted to remove any films, and

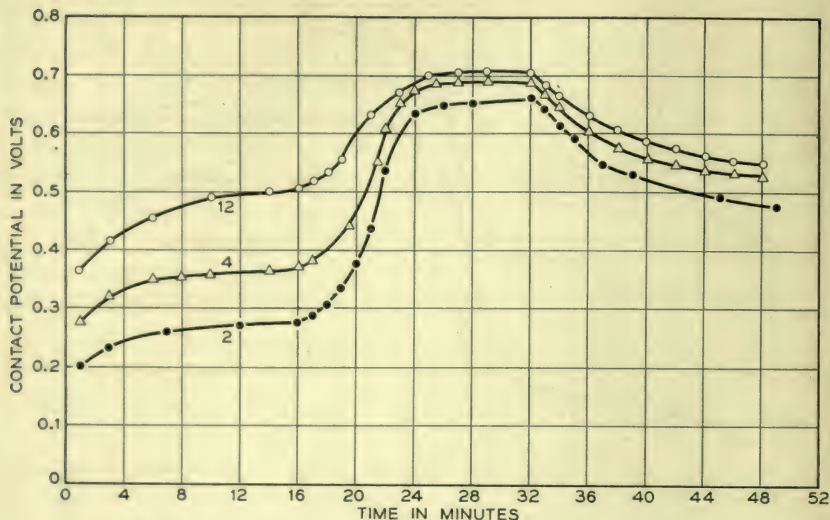


Fig. 4 — Contact potential cycles for etched sample D.

then given a polishing etch (CP-4). After the etch the surface was washed in running distilled water of reasonable quality. The surface was then dried with filter paper and kept covered until placed in the bell jar.⁴ From here on the procedure was the same as before. In this case cycles 1, 3 and 11 are shown. The surprising result is the similarity between Fig. 2 and Fig. 3. To a first approximation one is tempted to say that the dipole of a Ge surface, in the bell jar atmosphere is independent of past history. Within certain limits this is approximately true. There are differences between Fig. 2 and Fig. 3 but they are small and probably due to differences in surface treatment. Fig. 4 shows the results for *p*-type slice, D, when this surface is etched as above. This slice and the slice A were placed in the bell jar at the same time. Cycle 1, Fig. 3, was taken on A, cycle 2, Fig. 4, on D and so on. Any differences between the results in these two figures are to be attributed to the differences in samples. They cannot very well be ascribed to the reference electrode and the initial surface treatments were as nearly the same as they could be made with reasonable care. By making runs of this type two samples at a time, different samples and different surface treatments can be intercompared. This method eliminates the shifts in the work function of the reference electrode that sometimes occur from run to run. Such results can be illustrated by plotting the data for different samples and different surface treatments for cycles 10 or greater where the results for successive cycles are the same. This has been done in Fig.

5 where one cycle each is plotted for both sample A and D, for each of the two surface treatments, sandblasting and etching. The curves are approximately all of the same shape so that the differences between them can be described by giving the shift in contact potential necessary to superimpose the curves. This treatment works very well except for the first part of the cycle after the spark coil where the shift necessary is sometimes more and sometimes less than that needed to make all of two curves superimpose well. Note in Fig. 5 that the contact potential for sample A etched is always greater than for sample D etched. In the case of the sandblasted surface just the reverse is true. Also the contact potential for the etched surface of either sample is always larger than for the sandblasted.

Using these methods comparable results were obtained on two other *n*-type samples, C and E, of increasingly lower specific resistance. Taking advantage of the relation between carrier concentration and the position of the Fermi level we have plotted in Fig. 6, the contact potential for both the etched and the sandblasted surfaces versus the position of the Fermi level ($E_F - E_i$) in electron volts. The contact potential values used were taken from the saturation values in wet O_2 but the shapes of the curves would be much the same for any other point in the cycle. The contact potentials are thought to be accurate to about 0.01 volts. Solid lines have been drawn through the points. Note that the contact potential for the etched surface is always greater than for the sandblasted surface, i.e., the work function is always less and that this difference is greater when $E_F = E_i$ or the germanium is nearly intrinsic.

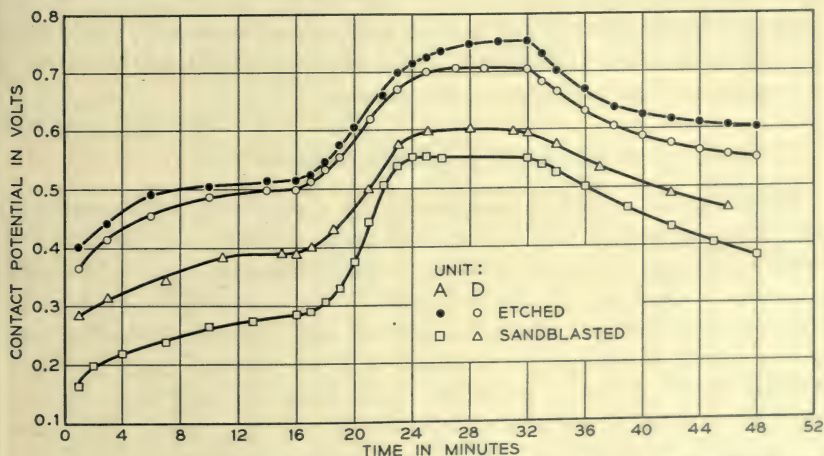


Fig. 5 — Comparison of cycles for samples A and D with sandblasted and etched surfaces.

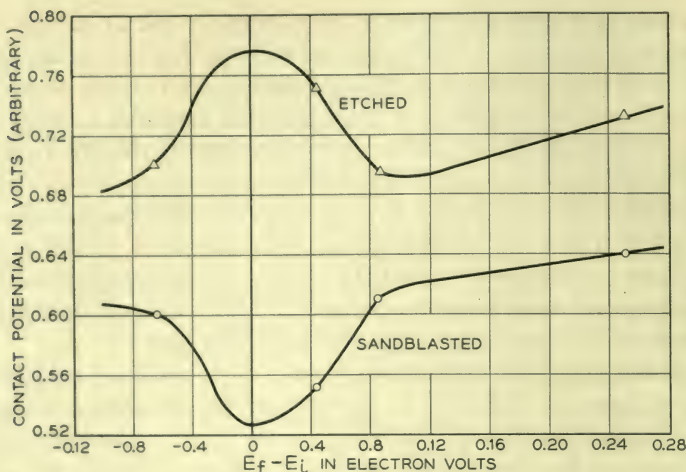


Fig. 6 — Contact potential dependence on position of the fermi level.

For variety Fig. 7 shows results on c.p. for a sandblasted polycrystalline sample of silicon. It is apparent that similar phenomena are taking place on the silicon surface, however the behavior is different. These preliminary results are shown to illustrate the generality of this method for investigating semiconductor surfaces in a controlled gaseous ambient.

The time between cycles was somewhat variable. The first and second cycles were always taken immediately after the specimen was placed in the bell jar. After this, successive cycles were taken one or two a day over a period of a week or more. The bell jar was not opened during a run and a small flow of dry gas was maintained between cycles. Little if anything occurred during these idle periods, showing that the changes that did take place were due to the cycling.

II. Change of contact potential with illumination

When the Ge surface is sandblasted the change of contact potential with illumination $(\Delta c.p.)_L$ is too small to be measured. If however the surface has been prepared by the polishing etch, the $(\Delta c.p.)_L$ is easily observed. This change if not too small can be measured by finding the balance on the potentiometer for light off and light on. When the contact potential is changing with time, or if the change is small this is difficult to do. A better procedure is to chop the light at a definite frequency and measure the amplified output on an ac meter. This gives a continuous reading that can be read easily at any given time. If a filter is used to pass only those frequencies near the fundamental of

the chopping frequency, the improvement in signal to noise enables one to read very small changes.

A plot of $(\Delta c.p.)_L$ in volts versus the contact potential for samples A *n*-type and D *p*-type is shown in Fig. 8. For the *n*-type sample the signal is large and positive when the contact potential is small and becomes small and negative as the contact potential increases. Except for the shift in the contact potential where $(\Delta c.p.)_L$ goes through zero, the results for the *p*-type sample are practically the opposite of those for the *n*-type sample. Similar curves have been obtained cycle after cycle in many complete runs. Within experimental error the curves have the same shape for a given sample for all cycles in all runs. Sometimes the curve for the first cycle in a run will differ in shape from the rest. However there are shifts such that the c.p. for which the light goes through zero $(c.p.)_0$ does vary from cycle to cycle throughout a run. Fig. 9 is a plot of $(c.p.)_0$ versus cycle number for *p*-type sample D and *n*-type sample A. The data are plotted for two distinct runs. Both of these units were in the bell jar when the measurements were taken. Consistent results of this kind give one confidence that the Pt reference electrode is staying constant.

These experimental results can be summarized as follows. All data for a given sample can be superimposed by shifts in contact potential scales for the different cycles. A plot of $(\Delta c.p.)_L$ versus c.p. — $(c.p.)_0$ can be represented by a single curve for each unit. Moreover all the curves for all units both *n*- and *p*-type have quite similar shapes provided in the latter we plot $[-(\Delta c.p.)_L]$ versus $(c.p.)_0 - c.p.$ The shape of this curve is shown in Fig. 8 for a given sample and Fig. 9 shows how $(c.p.)_0$ varies from cycle to cycle. These plots adequately describe the results.

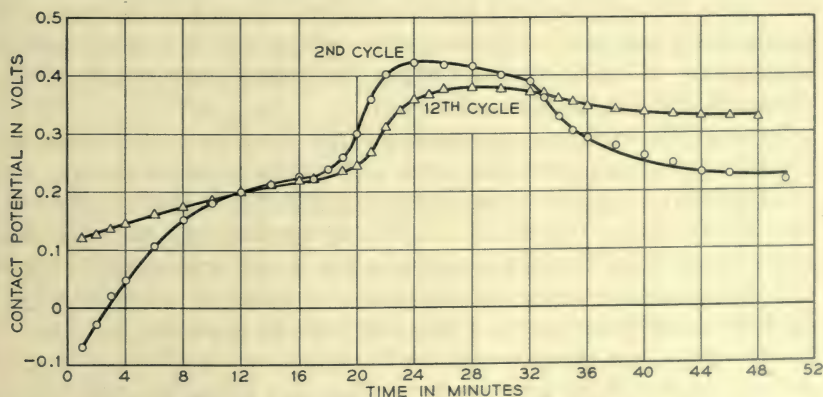


Fig. 7 — Contact potential cycles for silicon.

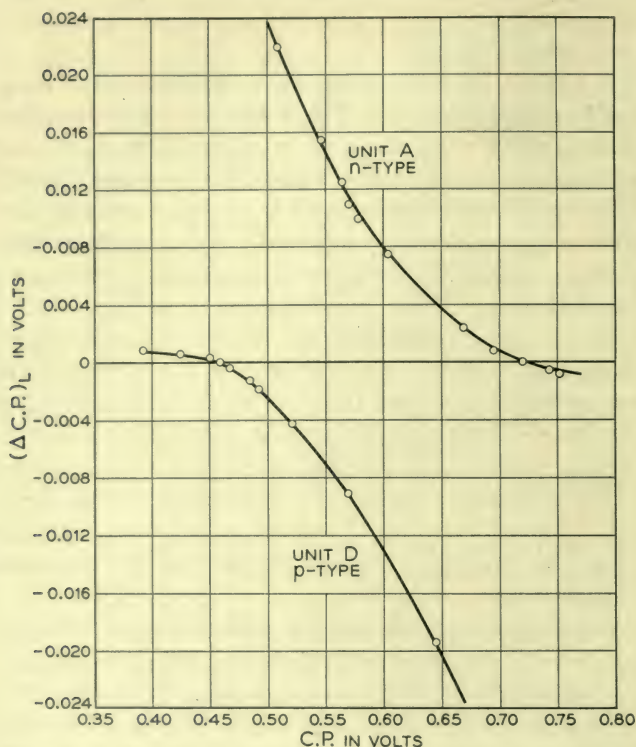


Fig. 8 — Change of contact potential with illumination versus contact potential, samples A and D etched.

Not shown in Fig. 8 because of the scale used is the result that $(\Delta c.p.)_L$ for *n*-type material does not increase indefinitely as c.p. decreases but approaches a maximum value. Likewise $-(\Delta c.p.)_L$ for *p*-type approaches a maximum as c.p. increases. Figures illustrating these results will be discussed after the theory is presented.

Some of the experimental details require discussion. It is necessary to calibrate the ac response in terms of absolute potential change. This can be done by comparing the ac reading with the dc reading when the light signal is large. One can also do this by introducing a known square-wave signal across the potentiometer in Fig. 1, and reading the ac signal out. Both methods agree when allowance is made for variation of the light signal with frequency. The latter response is almost flat from 25 to 300 cycles, but there is evidence for some very low frequency components in the dc measurements of $(\Delta c.p.)_L$. When the light signal goes through zero the signal is small and the dc value is difficult to read.

At times evidence has been obtained to indicate that the dc value changed sign. At other times the dc $(\Delta c.p.)_L$ behaved as if the place where it went through zero was shifted in c.p. from the point where the ac signal goes through zero. In view of this it was necessary to prove that the ac signal was changing phase at the zero point and not just going down in the noise and then increasing again without phase change. This was done by comparing the phase of the signal with a signal from a photocell placed in the same chopped light beam. By this means it was proved conclusively that the ac light signal was actually going through zero.

Some data were obtained on change of contact potential with light on *n*-type samples C and E having progressively smaller specific resistances. It was found that $(\Delta c.p.)_L$ decreased with specific resistance. It also decreased into the noise as the contact potential was increased, so that it could not be determined if it changed sign as for samples A and D. Because of the smaller signal $(\Delta c.p.)_L$ could not be measured easily except by the ac method and so far the signal has not been calibrated properly.

Some preliminary data on *p*-type silicon indicate that $(\Delta c.p.)_L$ for this sample was negative and that it decreased as c.p. was decreased. The magnitude of $(\Delta c.p.)_L$ for the same light intensity was much larger than for germanium and so far it has not been found to go through zero and change sign in the experimental range.

III. Other methods of varying the c.p.

In some cases N_2 was used in place of O_2 . A spark discharge in the N_2 had very little effect on the c.p. On the other hand wet N_2 produced much the same effect as wet O_2 . The positive extreme in c.p. was about 0.1 volt greater in the case of wet N_2 . After the wet treatment dry N_2 was not nearly as effective as dry O_2 in reducing the c.p. to its intermediate value. This can hardly be due to a difference in dryness of the two gases since the same drying column was used in both cases. The results indicate that dry N_2 tended to leave the surface in whatever condition obtained before the dry N_2 flow was started, and that O_2 counteracts the effect of H_2O .

With dry N_2 as a carrier, other vapors were tried. A. N. Holden suggested trying a peroxide and picked out ditertiary butyl peroxide as being reasonably safe. Use of this vapor was found to produce the same changes as the spark coil in the O_2 flow. Other vapors having OH radicals such as methyl alcohol and acetic acid were found to act the same way

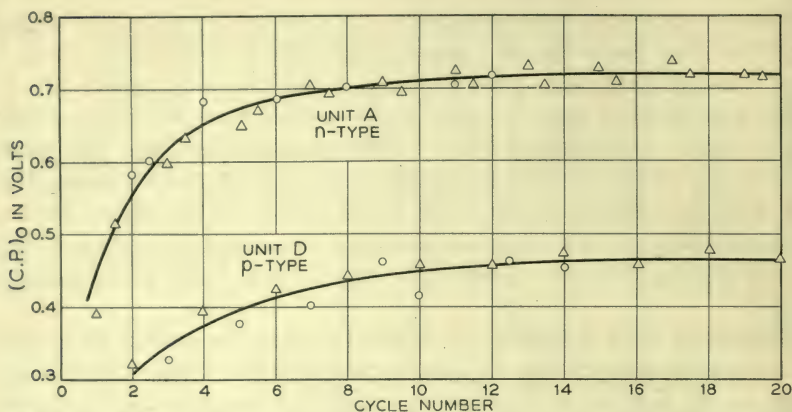


Fig. 9 — Contact potential for zero light effect, $(c.p.)_0$, versus cycle number, two runs for each sample.

as water vapor. To prove that this was not caused by traces of water in the alcohol or acid the N_2 was bubbled through water solutions of H_2SO_4 . These results indicated that one needs appreciable amounts of water vapor to produce the effect, much more than could be present in the alcohol or acid. Other vapors, such as carbon tetrachloride, methylchloride, nitrobenzene and ether, were found to have no effect on either the contact potential or $(\Delta c.p.)_L$. Acetone has a small effect in the same direction as water. This is to be expected because this compound exists in part in a tautomeric form having an OH group. Vapor from 30 per cent H_2O_2 , 70 per cent H_2O acted at first like a peroxide vapor and with a longer time of exposure behaved like water vapor. Small amounts of Cl_2 gas in N_2 produced the same change as the spark discharge in O_2 and after 14 minutes of flushing the bell jar with N_2 produced an additional effect when water vapor was introduced. On *n*-type samples before the usual increase in c.p. and decrease in $(\Delta c.p.)_L$ there was a large increase in $(\Delta c.p.)_L$. This was attributed to the reaction between the water vapor and the Cl left on the Ge surface, producing oxygen. The nature of the change was a rapid shift in $(\Delta c.p.)_0$ and thus a momentary increase in $(\Delta c.p.)_L$. This effect is only large in the first cycle. It indicates that the shifts in $(\Delta c.p.)_0$ plotted in Fig. 9 are probably due to an oxidation of the Ge surface as the cycling progresses. In all these experiments the relation between the c.p. and $(\Delta c.p.)_L$ was essentially the same as that obtained in the standard cycle.

One can detect the presence of thin surface films by electron diffraction techniques. R. D. Heidenreich took electron diffraction pictures of a germanium surface immediately after the polishing etch and water

wash. He found the surface to be quite clean, with a film thickness less than $10 \text{ \AA}$. He also took pictures after the germanium surface had been cycled about fifteen times and found in this case definite evidence for a thin surface film. The film was either amorphous or composed of very small crystals. He estimated the thickness to be between 20 and $50 \text{ \AA}$.

In the discussion of the experimental work it has been assumed that all the changes in c.p. are to be attributed to the Ge surface and not to the Pt. It is not easy to give a definite proof that this is true. The fact that it is a reasonable assumption is suggested by the nature of the results themselves. Almost identical results were obtained using a Ta electrode. In Fig. 10 we show the results of two cycles when the Ge was replaced with gold. While there are some changes they are almost an order of magnitude smaller than the changes when Ge is present. It follows that one would expect much the same results with either Ta, Pt or Au reference electrodes and that most of the changes are due to the Ge. No change of c.p. with illumination is observed when both electrodes are metals. In one case a small amount of H_2 was added to the N_2 flow. Here the c.p. between Pt and Ge changed rapidly but the $(\Delta \text{c.p.})_L$ did not change at all. Subsequent runs using the regular cycle indicated that the c.p. scale had been shifted corresponding to a reduction in work function of the Pt. Except for this shift the results were the same and the shift disappeared in about one day. The conclusion was that H_2 had little effect on Ge but reacted with the Pt decreasing its work function. This is a good illustration of the power of this method of measuring more than one property of a semiconductor surface at the same time. If only c.p. had been measured the conclusions would not have been so clear cut.

If one knew the work function of the Pt electrode then one would know the work function of the Ge. Work functions are all measured in high vacuum. We have not been able to think of a method of determining the work function of any electrode in a gaseous ambient unam-

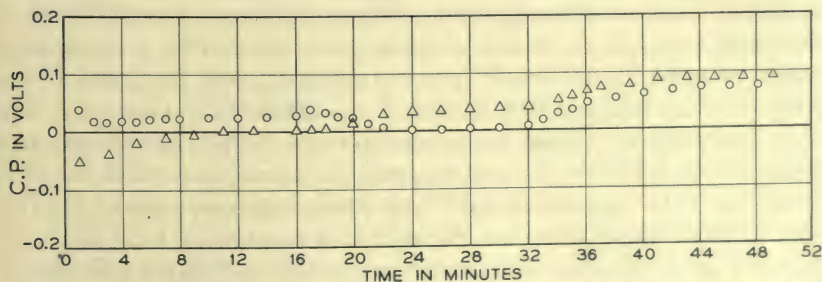


Fig. 10 — Contact potential cycles when germanium is replaced by gold.

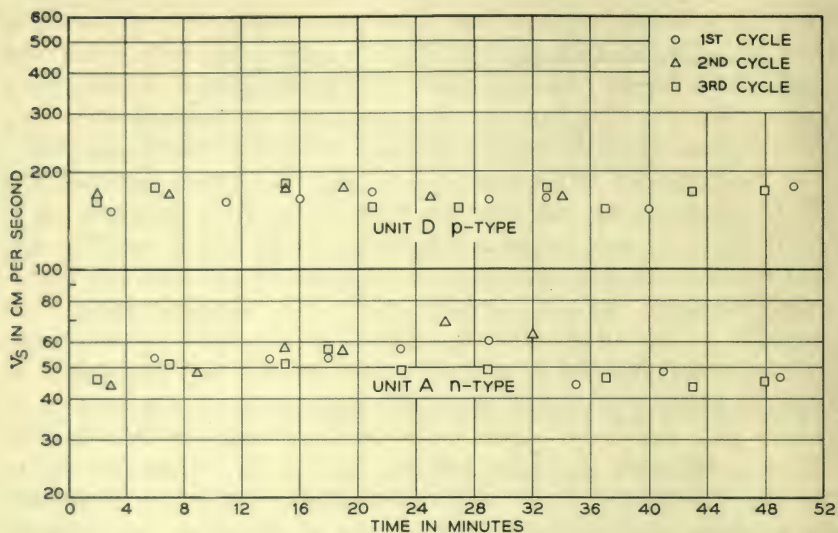


Fig. 11 — Dependence of the velocity of recombination, v_s , on the contact potential cycles.

biguously. It may well be impossible to do this in Bridgman's operational sense. Most physicists would agree that the Pt reference electrode probably has a work function of around 5 to 7 volts under these conditions, but this of course does not help much.

IV. Measurement of surface recombination velocity v_s

At first we tried various methods of measuring v_s on the Ge surface at the same time that c.p. and $(\Delta c.p.)_L$ were measured. No method was found that could be trusted. As the study progressed we realized that the c.p. of a Ge surface could be cycled in a reproducible way. Since the proper geometry for measuring v_s was a rod or filament,² rods were prepared of samples A and D with the same chemical surface treatment. The decay of hole electron pairs, created by a point light source, was measured as a function of distance from the light. If the body life time τ and the dimensions of the filament are known one can then determine v_s .² These measurements were made while the gaseous ambient was cycled in the same manner as before. The results for filaments cut from samples A and D are plotted against time in Fig. 11. The ambient atmosphere was changed as a function of time just as it was in Figs. 2 to 4. The main conclusion is that within the accuracy of this experiment there is no evident dependence of v_s on the changes in

gas ambient and therefore no dependence on the corresponding changes in surface dipole. This result was somewhat unexpected and the first time the experiment was performed it was hard to believe that the previously measured changes in c.p. actually were taking place. In this case the gas ambient was experimented with to try to change v_s and it was found that v_s could be changed from the order of 10^2 cm/sec to greater than 10^5 cm/sec and back again by exposure of the filament to $(\text{NH}_4)\text{OH}$ fumes and then HCl fumes respectively.⁵ While interesting, this has no direct bearing on the other experiments. The experiment was performed again with freshly prepared rods and this time the cycling used in the c.p. experiments was rigidly adhered to, giving v_s equal to 70 cm/sec and 200 cm/sec approximately for samples A and D respectively. The experiments were then repeated again using new filaments cut from the samples as close as possible to the surfaces used in the c.p. experiments leading to the results shown in Fig. 11, namely, v_s equal to 50 cm/sec and 170 cm/sec respectively. From these experiments it was concluded that v_s is approximately constant in the range involved and is determined by the nature of the sample and the surface treatment used. It was noted in some of these experiments that v_s for the first cycle was somewhat larger than for the subsequent cycles. This change when it occurred is probably to be correlated with the changes in the early cycles in the c.p. measurements. Similar measurements on sample C gave v_s equal to 1500 cm/sec. No measurements of v_s were made on samples B and E.

V. Other experimental measurements

The specific resistance of each sample was measured near the surface used in the experiments. It was approximately constant across the surface but did vary slowly with depth in some of the samples.

The body life times were measured on each sample. The thickness of the slices used was intentionally made large compared to their corresponding diffusion lengths, about 0.5 cm for A, B, C and E and about 2.0 cm for D. The mobilities were taken from J. R. Haynes³ measurements: $\mu_n = 3600$ and $\mu_p = 1700$ cm²/volt sec. There is some uncertainty as regards the exact value of the equilibrium product of holes and electrons, np , at 300°K. We have used the value 6.3×10^{26} obtained from some unpublished data of G. L. Pearson.³

The light source was calibrated by replacing the germanium sample with one of F. S. Goucher's n - p junctions.⁶ The bell jar, with everything else including the reference electrode, was left in their normal

positions. An average was taken over the filament image, and the effective area of the p - n junction was determined and allowed for. This was done for all light intensities used. Most of the experiments were performed with a fixed intensity and the averaged result for this intensity was 6.0×10^{15} , hole electron pairs per cm^2 sec. The rate of pair production was found to be proportional to the light intensity.

Since practically all the light is absorbed in a depth (10^{-4} cm or less) that is small compared to the diffusion length, it is a simple matter to calculate the steady state increase δp in hole electron pairs due to the light. The relation is

$$\delta p = N/(v_s + v_d) \quad (1)$$

where N is the rate of pair production, v_s is the velocity of recombination at the surface and v_d is the diffusion velocity for the minority carrier. Since N is proportional to light intensity it follows that δp is too.

The magnitude of $(\Delta c.p.)_L$ should depend on the light intensity. One might at first expect it to be proportional to light intensity. That this is not the case is shown by curve 1 in Fig. 12. Curve 1 is a plot of $(\Delta c.p.)_L$ versus δp for unit D on a log-log scale. A smooth curve has been drawn through the experimental points. As we shall see later, theory predicts that if $(\Delta c.p.)_L$ is large, it should be proportional to $\ln(1 + \delta p/a)$ where a is the equilibrium density of the minority carrier,

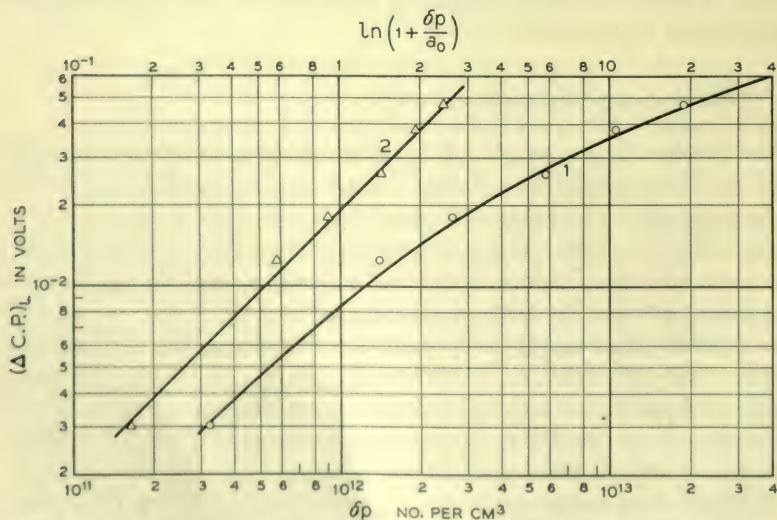


Fig. 12 — Dependence of contact potential change with illumination $(\Delta c.p.)_L$ on light intensity.

n in p -type material and p in n -type. That this prediction is borne out is shown by how well the experimental points fit the straight line curve 2 in Fig. 12 where $(\Delta c.p.)_L$ is plotted versus this quantity. The scale for δp is shown along the bottom of Fig. 12 and that for $\ln(1 + \delta p/a)$ along the top. Similar results were obtained for unit A.

In Table I the parameters, specific resistance ρ in ohm cm, life time τ in microseconds and the surface recombination velocity v_s in cm/sec are given for each unit used. Also given are some pertinent quantities derived therefrom, namely the equilibrium densities of electrons and holes, n and p in number per cm^3 and the increase in density at the surface δp in number per cm^3 when the rate of pair production due to the illumination was 6.0×10^{15} per cm^2 sec.

THEORY

The constancy of v_s throughout the range of surface dipole investigated puts rather stringent requirements on any theoretical model to be constructed. W. Shockley and W. T. Read⁷ have investigated the theory of recombination via traps. It is evident from their work that if one assumes a trap density peaked near a single energy and very small elsewhere, then v_s will be constant over a range of surface dipole values provided that the peak energy is either near the conduction band or near the filled band. The experimental results make it appear very unlikely that the trapping mechanisms on the n and p -type surfaces are essentially different. It is assumed that both types of traps are present on the surface and that the traps are approximately the same for both n - and p -type samples. Further, it is assumed that the traps N_a of energy E_a near the conduction band are donor type, i.e., neutral when filled and positively charged when empty. Likewise the traps N_b of energy E_b near the filled band are assumed to be acceptor-type traps, i.e., neutral when empty and negatively charged when filled. The absolute charge on the traps is not important, however, because we are

TABLE I

Sample	Type	ρ	n_0	p_0	τ	v_s	δp	C
A	n	12.5	1.38×10^{14}	4.56×10^{12}	900	50-70	2.2×10^{13}	4.4×10^{-13}
B	n	15	1.14×10^{14}	5.6×10^{12}	600	(100)	1.9×10^{13}	—
C	p	12.0	2.1×10^{12}	3.0×10^{14}	4000	170-200	1.75×10^{13}	6.0×10^{-13}
D	n	2.5	7.0×10^{14}	0.91×10^{12}	48	1.5×10^3	1.67×10^{12}	20.0×10^{-13}
E	n	0.008	4.4×10^{17}	1.45×10^9	—	—	—	—

concerned only with differences between the occupied and unoccupied states.

The fraction of the traps which are occupied depends on the trap energy and on the position of the Fermi level at the surface. The latter in turn depends on the condition that the surface as a whole be neutral. A very obvious mechanism for changing the total surface dipole is the adsorption or desorption of ions on the surface. If this happens the position of the Fermi level at the surface shifts until the total charge on the surface, adsorbed ions, charge in the traps and charge in space-charge layer, adds up to zero. The consequences of this model have been carried through.

The basis for the theory is illustrated in the energy level diagram of Fig. 13. At the semiconductor surface there is a space-charge layer of thickness ℓ_B which gives a change in electrostatic potential of V_B , corresponding to a potential energy of an electron of $-eV_B$. Outside of the surface of the germanium proper, there is a surface film of thickness ℓ_D . A double layer giving a potential change V_D , is formed from a charge of ions, σ_I , on the outer surface of the film and charges in the surface traps of types *a* and *b*. Changes in c.p. with ambient result from changes in σ_I and consequent changes in V_B and V_D . It is assumed that the remainder of the work function is independent of ambi-

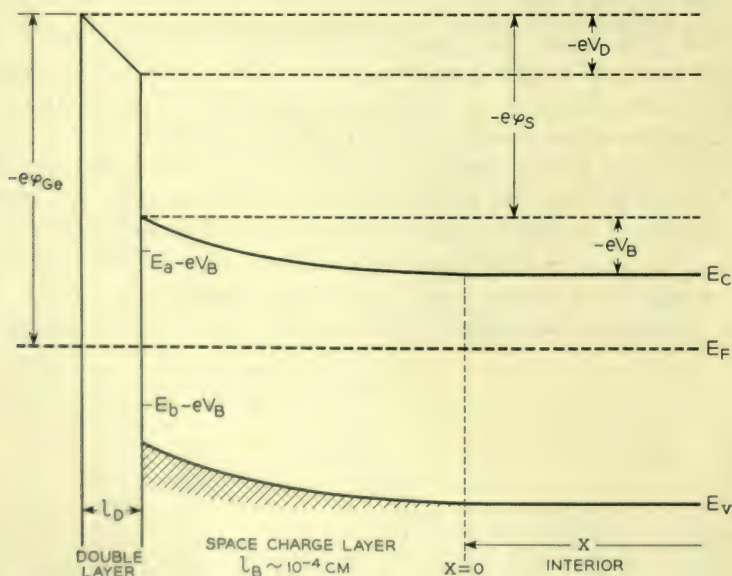


Fig. 13 — Schematic of energy level diagram at germanium surface.

ent. When light shines on the surface, V_B is changed to $V_B + \delta V_B$ and there is an additional potential drop, δV_i , in the body of the germanium resulting from the recombination current which flows to the interior. The change in contact potential with light is equal to $\delta V_B + \delta V_i$.

The film thickness ℓ_D is shown on an exaggerated scale. We expect $\ell_D \sim 10^{-6}$ cm and $\ell_B \sim 10^{-4}$ cm, so that $\ell_B \gg \ell_D$.

TABLE OF SYMBOLS

A. Energies:

E_a = E_a (true) + $kT \ln (\omega_{unoc}/\omega_{oc})$, is the effective energy of the a -traps for $V_B = 0$. Here ω_{unoc} and ω_{oc} are the statistical weights of the unoccupied and occupied states, respectively.

E_b = effective energy of the b -traps for $V_B = 0$.

E_c = energy of lowest state of conduction band in interior of semiconductor just beyond the space-charge layer.

E_v = energy of highest state of valence band at the same position.

E_F = Fermi energy.

E_i = E_F when material is intrinsic.

V_D = potential drop across surface film.

V_B = potential drop across space-charge layer.

V_{B0} = value of V_B for which $n_a = p_b$, see below.

V_0 = value of V_{B0} for an intrinsic sample.

B. Concentrations:

n = $N_c \exp [(E_F - E_c)/kT]$ = equilibrium concentration (no./cm³) of conduction electrons in interior of semiconductor just beyond the space-charge layer.

p = $N_v \exp [(E_v - E_F)/kT]$ = corresponding hole concentration.

n_i = intrinsic concentration.

n_s, p_s = equilibrium concentrations of electrons and holes, respectively, at the surface.

N_a, N_b = concentration (no./cm²) of a - and b -traps, respectively.

n_a = equilibrium concentration (no./cm²) of occupied a -traps.

p_b = $N_b - n_b$ = equilibrium concentration (no./cm²) of unoccupied b -traps.

n_{a0}, p_{b0} = values of n_a and p_b for an intrinsic sample with $V_B = 0$.

n_1 = $n + \delta n$, $p_1 = p + \delta p$, and $n_{s1}, p_{s1}, n_{a1}, p_{b1}$ = concentrations in presence of light. Electrical neutrality requires that $\delta n = \delta p$.

The theory is based on the following postulates:

I. Changes in c.p. with ambient result from changes in σ_I and the consequent changes in V_B and V_D .

$$\text{c.p.} = V_B + V_D + \text{const.} \quad (2)$$

The charge σ_I is largely compensated by charges in the surface traps. The barrier height, V_B , is determined by the requirement of electrical neutrality:

$$\sigma_I = e(n_a - p_b) + \text{const.} \quad (3)$$

Since V_B and V_D are of the same order, the net charge per unit area in the space-charge layer will be smaller than σ_I in the approximate ratio ℓ_D/ℓ_B and may be neglected.

II. Traps of type a have energies above E_F and of type b below E_F for all values of V_B attained in the different ambients used. More exactly

$$E_a - eV_B - E_F > kT, \quad (4)$$

$$E_F - E_b + eV_B > kT, \quad (5)$$

for all V_B . One may then use the Boltzmann approximations for n_a and p_b :

$$n_a = \frac{N_a}{1 + \exp [(E_a - eV_B - E_F)/kT]} \quad (6)$$

$$\sim N_a \exp [(E_F - E_a + eV_B)/kT],$$

$$p_b = \frac{N_b}{1 + \exp [(E_F - E_b + eV_B)/kT]} \quad (7)$$

$$\sim N_b \exp [(E_b - eV_B - E_F)/kT].$$

It is not necessary for our arguments to assume that all traps of each type have the same energy. The only requirement is that the distributions of trap energies are such that the a -traps are always above and the b -traps always below the Fermi level for any ambient.

III. Creation of electron-hole pairs by absorption of light occurs near the surface in a distance that is small compared with the diffusion length. Optical constants of germanium indicate that practically all of the light with energy sufficient to create electron-hole pairs is absorbed within a distance of 10^{-4} cm of the surface. The diffusion length is of the order of 0.2 cm.

IV. In the presence of light, the concentration of electrons in a -traps

is in equilibrium with the concentration of electrons in the conduction band and holes in b -traps are in equilibrium with holes in the valence band. The barrier height is adjusted so that the total charge in both types of traps is unchanged by illumination. We shall show in the appendix that the resistance to flow of electrons from the conduction band across the space-charge layer and into a -traps is small compared with the resistance to flow of electrons from the valence band to the traps. Similar considerations apply to flow of holes to b -traps from the valence band as compared with flow from the conduction band.

V. Recombination is limited by holes going into traps of type a and electrons going into traps of type b . The two types of traps act in parallel for recombination. The contributions to the surface recombination velocity are proportional to $p_s n_a$ and $n_s p_b$, respectively. These products are independent of V_B and thus of ambient if postulates II and IV are satisfied. If other types of traps were important in recombination, one would expect v_s to depend on ambient, contrary to what is observed.

Postulates I and II are used to relate V_B and c.p. with changes in ambient. Postulates III, IV, and V are used to relate $(\Delta c.p.)_L$ with V_B and the trap densities, and also to obtain an expression for the surface recombination velocity.

As V_B is made more positive, corresponding to a decrease in barrier height for electrons, n_a increases and p_b decreases. It will be convenient for the theoretical discussion to introduce the particular barrier potential V_{BO} , for which $n_a = p_b$. With use of the Boltzmann approximations in Equations (6) and (7), this gives

$$\begin{aligned} \exp [2\beta V_{BO}] &= (N_b/N_a) \exp [(E_a + E_b - 2E_F)/kT] \\ &= (p_{b0}/n_{a0})(p/n), \end{aligned} \quad (8)$$

where $\beta = e/kT$. The last form follows from the definitions of p_{b0} and n_{a0} and by noting that $\exp [2(E_i - E_F)/kT] = p/n$. We then have

$$n_a/p_b = \exp [2\beta(V_B - V_{BO})]. \quad (9)$$

As so defined, V_{BO} depends on the Fermi level and thus on the conductivity of the specimen. We shall let V_0 be the value of V_{BO} for an intrinsic specimen. From (8)

$$n_{a0}/p_{b0} = \exp [-2\beta V_0]. \quad (10)$$

If $n_{a0} = p_{b0}$, then V_0 will be zero. Postulate II sets limits on V_0 , but V_0 is otherwise undetermined in our experiments.

We shall first relate V_D and V_B . Since the electric field in the film is $4\pi\sigma_I/K_D$,

$$V_D = 4\pi\ell_D\sigma_I/K_D, \quad (11)$$

where K_D is the dielectric constant of the film. Substituting for σ_I from Equation (3) and making use of (8) and (9), we find

$$V_D = 2H \sinh \beta(V_B - V_{Bo}) + \text{const}, \quad (12)$$

where

$$H = \frac{4\pi e\ell_D}{K_D} (N_a N_b \exp [(E_b - E_a)/kT])^{\frac{1}{2}} = (4\pi e\ell_D/K_D)(n_{a0}p_{b0})^{\frac{1}{2}}. \quad (13)$$

If V_D is expressed in volts and ℓ_D in cm.

$$H = 1.8 \times 10^{-6} (\ell_D/K_D)(n_{a0}p_{b0})^{\frac{1}{2}}. \quad (13a)$$

The contact potential Equation (2) may be expressed in the form

$$\text{c.p.} = V_B - V_{Bo} + 2H \sinh \beta(V_B - V_{Bo}) + \text{const}. \quad (14)$$

The change in contact potential with illumination results from a potential drop δV_i in the interior and a drop δV_B across the space-charge layer. The former comes from the recombination current of holes and electrons diffusing from the surface to the interior. Since electrical neutrality requires that $\delta n = \delta p$, the concentration gradients are equal. However, the mobility of electrons is greater than that of holes, so that the diffusion current of electrons is larger than that of holes by the mobility ratio "b." Since there can be no net current flow to the interior, an electric field is established which is in such a direction as to enhance the flow of holes and retard the flow of electrons. The net potential drop associated with this electric field is

$$\delta V_i = \frac{(b-1)}{\beta(b+1)} \ln \left\{ 1 + \frac{(b+1)\delta p}{bn+p} \right\} \quad (15)$$

where $\delta p = \delta n$ is the change in concentration in the interior just beyond the space-charge layer.³

This potential is positive for both *n*- and *p*-type material, and is independent of ambient, since δp is independent of ambient. The observed $(\Delta \text{c.p.})_L$ is generally much larger than given by (15) and is of opposite sign for *n*- and *p*-type material. To account for the observations, it is necessary to assume that the major part of the effect is associated with a space-charge layer at the free surface. We believe that the present experiments give the most convincing evidence obtained so far for the existence of such a space-charge layer.¹

The value of δV_B , the change in potential across the space-charge layer due to light, is determined by the requirement that there be no net change in charge in the surface traps, or that

$$\delta n_a = \delta p_b. \quad (16)$$

The changes δn_a and δp_b come both from changes in δp and δn and from δV_B . According to postulate IV, $n_{a1} = n_a + \delta n_a$ is in equilibrium with the conduction band and $p_{b1} = p_b + \delta p_b$ with the valence band.

We have then

$$n_{a1}/n_a = n_{s1}/n_s = (n_1/n) \exp [\beta \delta V_B], \quad (17)$$

$$p_{b1}/p_b = p_{s1}/p_s = (p_1/p) \exp [-\beta \delta V_B]. \quad (18)$$

from which it follows that

$$\delta n_a/n_a = (n_1/n) \exp [\beta \delta V_B] - 1, \quad (19)$$

$$\delta p_b/p_b = (p_1/p) \exp [-\beta \delta V_B] - 1, \quad (20)$$

$$\frac{n_a}{p_b} = \exp [2\beta(V_B - V_{BO})] = \frac{(p_1/p) \exp [-\beta \delta V_B] - 1}{(n_1/n) \exp [\beta \delta V_B] - 1}. \quad (21)$$

Equation (21) is a quadratic equation in $\exp [\beta \delta V_B]$ which may be solved to give an explicit expression for δV_B . The role of electrons and holes may be interchanged by changing the sign of δV_B and of $V_B - V_{BO}$. This accounts for the difference in behavior of n - and p -type samples. The total change with light is the sum of δV_i and δV_B :

$$(\Delta c.p.)_L = \delta V_L = \delta V_B + \delta V_i. \quad (22)$$

In the analysis of the data, δV_i is calculated theoretically from (15) with $\delta p = \delta n$ determined from (1) and δV_B is obtained from the observed $(\Delta c.p.)_L$ and δV_i using (22). Equation (21) is then used to find $V_B - V_{BO}$. A plot of $c.p. - (V_B - V_{BO})$ versus $\sinh \beta(V_B - V_{BO})$ should be a straight line with slope $2H$. An analysis of the observed data in this manner is given in the following section.

We turn finally to a discussion of the surface recombination velocity, v_s . According to postulate V, recombination is limited by flow of holes to a -traps and of conduction electrons to b -traps. The flow of holes from the valence band to a -traps (really electrons from a -traps drop into the vacant levels in the valence band corresponding to the holes) is proportional to the product of the hole concentration at the surface p_{s1} , and the concentration of electrons in a -traps, n_{a1} . The reverse flow is that of thermal generation of holes: electrons from the valence

band go into unoccupied a -traps. Since, according to postulate II, the number of unoccupied traps is nearly equal to the total number of traps, and is thus independent of δV_B and δp , the reverse flow will be practically equal to the thermal equilibrium value. If S_{pa} is the recombination cross-section for holes, the net flow of holes to a -traps is

$$U_{pa} = S_{pa}v_p(p_{s1}n_{a1} - p_s n_a)(\text{holes/cm}^2 \text{ sec}). \quad (23)$$

Here, v_p is the velocity factor which when multiplied by the concentration gives the number of holes crossing a unit area from one direction:

$$v_p = (kT/2\pi m_p)^{\frac{1}{2}}. \quad (24)$$

With use of the relations, $n_{a1}/n_a = n_{s1}/n_s$, $p_{b1}/p_b = p_{s1}/p_s$, from equations (17) and (18), which follow from the fact that the a -traps are in thermal equilibrium with the conduction band and the b -traps with the valence band, (23) becomes

$$U_{pa} = S_{pa}v_p(n_a/n_s)(p_{s1}n_{s1} - p_s n_s). \quad (25)$$

This expression may be simplified further. The ratio

$$\frac{n_a}{n_s} = \frac{N_a \exp [(E_F - E_a - eV_B)/kT]}{N_c \exp [(E_F - E_c - eV_B)/kT]} = \frac{N_a}{N_c} \exp \left(\frac{E_c - E_a}{kT} \right) \quad (26)$$

is independent of E_F and V_B . The ratio may be evaluated for an intrinsic specimen with $V_B = 0$, in which case $n_a = n_{a0}$ and $n_s = n_i$. Thus

$$n_a/n_s = n_{a0}/n_i. \quad (27)$$

We also have from postulate IV,

$$p_{s1}n_{s1} = p_1n_1. \quad (28)$$

The equilibrium products are

$$p_s n_s = pn = n_i^2. \quad (29)$$

With use of (27), (28) and (29), (25) becomes

$$U_{pa} = S_{pa}v_p(n_{a0}/n_i)(p_1n_1 - pn). \quad (30)$$

Similarly, it is found that net flow of electrons to b -traps is equal to:

$$U_{nb} = S_{nb}v_n(p_{b0}/n_i)(p_1n_1 - pn). \quad (31)$$

The total rate of recombination is given by the sum of U_{nb} and U_{pa} and is given by an expression of the form:

$$U = U_{nb} + U_{pa} = C(p_1n_1 - pn) = C(n + p)\delta p. \quad (32)$$

It should be noted that the coefficient C is independent of the Fermi level and thus of the conductivity of the specimen, whereas v_s is not. The relation between them is:

$$\text{For } n\text{-type} \quad C = v_s/n, \quad (33)$$

$$\text{For } p\text{-type} \quad C = v_s/p. \quad (34)$$

Values of C may be determined empirically from observed values of v_s . Referring to Table I, for sample A, $v_s = 60$ cm/sec and $n = 1.4 \times 10^{14}/\text{cm}^3$, so that $C = 4.3 \times 10^{-13}$ cm⁴/sec. For sample D, $v_s = 180$ cm/sec and $p = 3.0 \times 10^{14}/\text{cm}^3$, so that $C = 6 \times 10^{-13}$ cm⁴/sec. The values of C are approximately the same, indicating that the traps are not much different for the two specimens.

The theoretical value of C involves n_{a0} and p_{b0} . It is the product $n_a p_b = n_{a0} p_{b0}$, which is related to the parameter H and which can be estimated from empirical data. To obtain the concentrations themselves, a value must be assumed for V_0 . We have

$$n_{a0} = (n_{a0} p_{b0})^{\frac{1}{2}} \exp [-\beta V_0], \quad (35)$$

$$p_{b0} = (n_{a0} p_{b0})^{\frac{1}{2}} \exp [\beta V_0]. \quad (36)$$

As we have mentioned previously, there is no way to determine V_0 from our experiments, although postulate II sets limits on its value.

Let us for simplicity assume that $S_{nb}v_n = S_{pa}v_p = S_t v$. Then, using (35) and (36) in (30) and (32), we have

$$C = 2S_t v (n_{a0} p_{b0} / n_i^2)^{\frac{1}{2}} \cosh \beta V_0. \quad (37)$$

This equation may be used to estimate the trapping cross-section S_t for an assumed V_0 .

COMPARISON BETWEEN THEORY AND EXPERIMENT

In Fig. 14 we have plotted c.p. - (c.p.)₀ - ($V_B - V_{BO}$) versus $2 \sinh \beta(V_B - V_{BO})$ for p -type sample D, (see Table I). Each symbol represents experimental points for one cycle. Results are shown for five different cycles not all in the same run. These results are typical of all the data obtained for this sample. It is seen that these data can be fitted over most of the range by a straight line as drawn, giving a value of $H = 0.02$ e.v.

In Fig. 15 we have used the same experimental results to plot δV_L versus c.p. - (c.p.)₀. Since the experimental values of δV_L cover a range of a factor of 20 or more, a logarithmic scale was used to plot

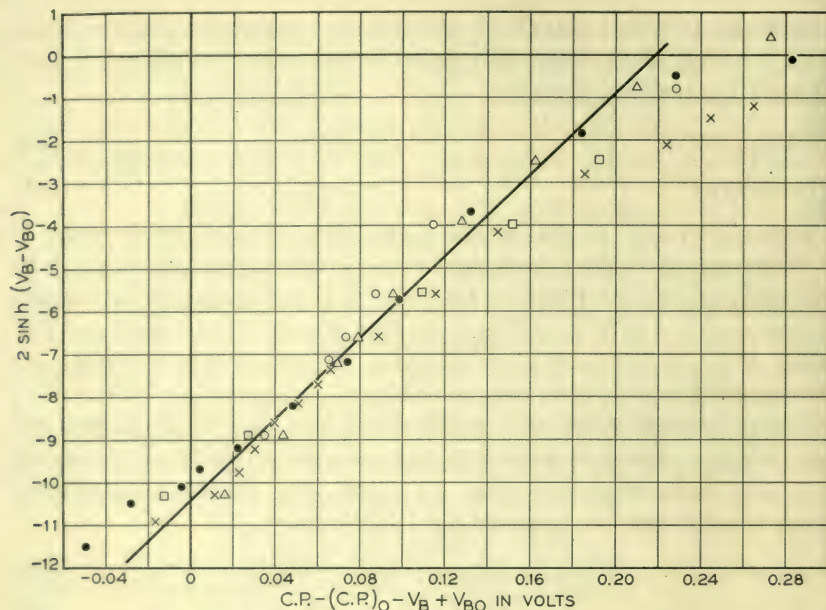


Fig. 14 — Plot of $2 \sinh \beta(V_B - V_{BO})$ versus c.p. $-(V_B - V_{BO})$ for sample D as suggested by theory.

δV_L . Both positive and negative branches of the curve are plotted on the same figure. The symbols used for the experimental points are consistent with Fig. 14. When c.p. $-(c.p.)_0$ is greater than zero, δV_L is negative. As c.p. increases it approaches a maximum negative value, this is the negative branch. When c.p. $-(c.p.)_0$ is less than zero, δV_L is positive and as c.p. decreases δV_L approaches a positive maximum value that is less in magnitude than the negative maximum. The solid curve represents the prediction of theory for $H = 0.02$ e.v. The agreement between theory and experiment is good. It should be emphasized that this fit is obtained with only one adjustable parameter.

The data for the other samples were analyzed in the same way. The results are shown by plotting δV_L versus c.p. $-(c.p.)_0$ as in Fig. 15. Fig. 16 is for n -type sample A and Fig. 17 for n -type sample B. The values obtained for H were 0.015 and 0.022 e.v. respectively. The fits obtained are about equally good in all cases with some deviation between theory and experiment near the extremes of contact potential. The values of H obtained are all of the same order as they should be if the surface trap structure is approximately the same from sample to sample. When $V_B - V_{BO}$ is large and positive and thus c.p., Eq. (14), is large, δV_L approaches the negative maximum

$$-\left[\left(\frac{1}{\beta} \ln \frac{n_1}{n}\right) - \delta V_i\right],$$

equation 21, and as $V_B - V_{BO}$ becomes small and negative, c.p. decreases and δV_L goes through zero and approaches the positive maximum

$$+\left[\left(\frac{1}{\beta} \ln \frac{p_1}{p}\right) + \delta V_i\right].$$

For p -type germanium n_1/n is greater than p_1/p so that the negative maximum for δV_L is greater than the positive maximum. Just the opposite is true for n -type germanium.

In this comparison of theory and experiment a tacit assumption has been made that the germanium surface is uniform in its properties. It might well be that this is not the case. The surface might be "patchy." To estimate what effect patches or non-uniformities might have the

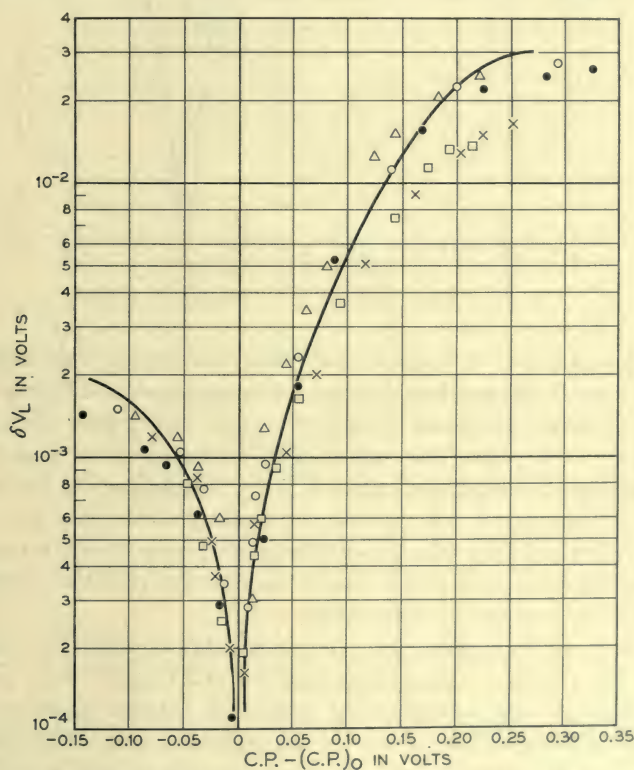


Fig. 15 — Change of contact potential with illumination, δV_L , versus c.p.; experiment and theory sample D, p -type.

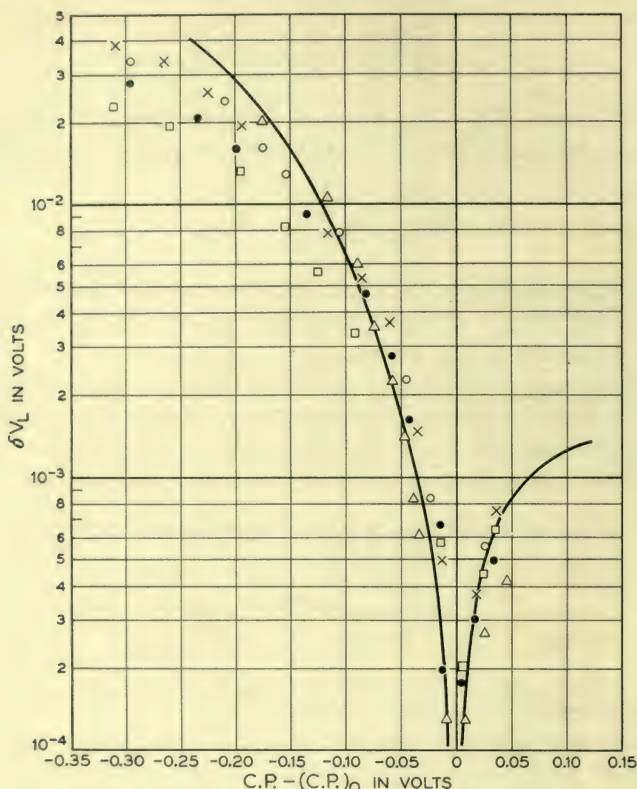


Fig. 16 — Same as Fig. 15 for sample A, *n*-type.

following was done. It was assumed that the surface was made up of two parts each having the theoretical dependence of δV_L on c.p. but with the contact potential where the light effect goes through zero differing by 0.05 volts. The values of δV_L at a given contact potential were averaged and plotted against c.p. The difference between this curve and the original one was almost entirely a simple shift in $(c.p.)_0$. It is therefore unlikely that non-uniformities in c.p., over the surface, of the order of 0.1 volts or less would change the relation between δV_L and c.p. sufficiently to be detectable.

One can use equations (8), (10) and (21) to calculate values of $(V_B - V_0)$ for each experimental value of δV_L and then plot these values against the corresponding values of contact potential. These results are shown in Fig. 18 for *n*-type sample A and *p*-type sample D. $(V_B - V_0)$ changes with c.p. as one would expect and in an approxi-

mately linear fashion throughout most of the experimental range. The change in V_B is approximately one-fifth the change in contact potential. ($V_B - V_0$) is positive for the p -type sample and negative for the n -type sample throughout most of the range. If the trap distribution on the surface were symmetrical about the intrinsic position of the Fermi level, then p_{b0} would be equal to n_{a0} and V_0 would be zero. In this case the space charge layers on n - and p -type germanium would be about equally developed. All the experimental evidence on germanium indicates that this is not the case but rather that $-V_{Br}$ is much larger than V_{Bp} . This then means that V_0 is negative and that the trap distribution is unsymmetrical either in number or energy or both in such a way that $n_{a0} > p_{b0}$.

In Table II values are given for the differences ($V_B - V_{B0}$), etc.,

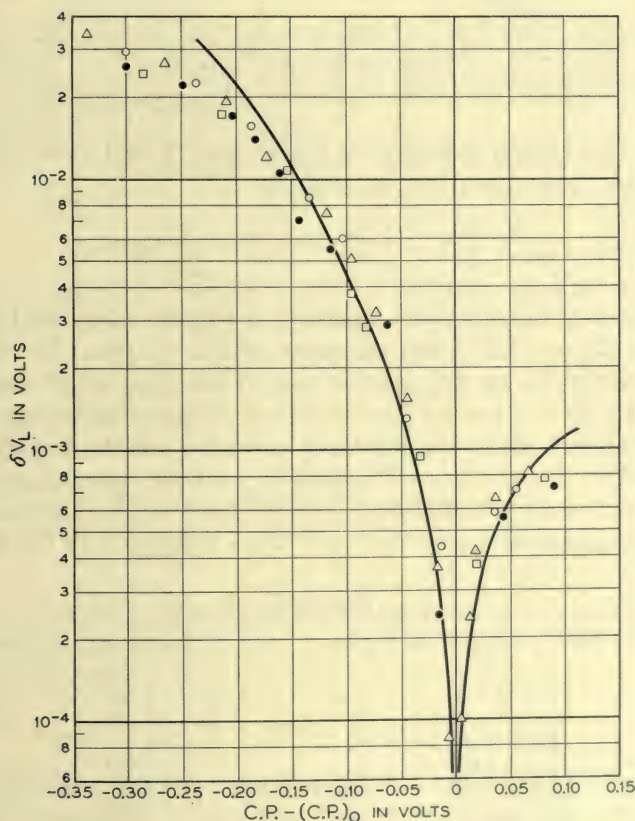


Fig. 17 — Same as Fig. 15 for sample B, n -type.

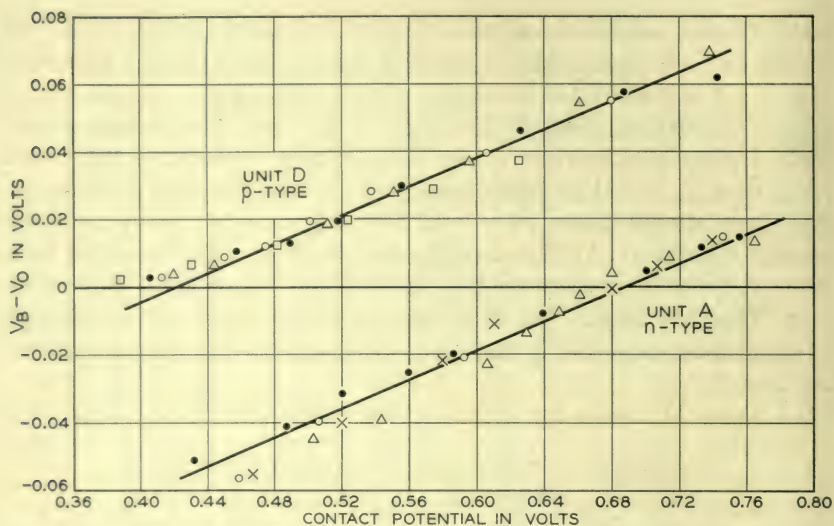


Fig. 18 — Potential across space charge layer, V_B , versus contact potential.

also δV 's, for certain special cases for samples A and D as calculated from theory. From equations (8) and (10)

$$(V_{BO} - V_0) = \frac{1}{e} (E_i - E_F) = \frac{1}{2\beta} \ln(p/n). \quad (38)$$

It is a constant for any given sample as is δV_i (see equation 15). From equations (21) and (38) it follows that when $\delta V_B = 0$ then $(V_B - V_0) = 0$ for all samples. To say this another way, if the traps are symmetrically distributed about E_i and V_B is zero, then illuminating the germanium would produce no potential difference across the surface. One would not suspect offhand that $(V_B - V_0)$ for $\delta V_L = 0$ was also approximately independent of the material in the sample but it is. For the case where δp and the minority carrier are both small compared to the majority

TABLE II

	A (n-type)	D (p-type)
$(V_{BO} - V_0)$	-0.044	+0.064
δV_i	+0.0019	+0.0015
$(V_B - V_0)\delta V_B = 0$	0	0
$(V_B - V_{BO})\delta V_B = 0$	+0.044	-0.064
$(V_B - V_0)\delta V_L = 0$	+0.010	+0.010
$(V_B - V_{BO})\delta V_L = 0$	+0.054	-0.054
$\delta V_L, (V_B - V_{BO}) = 0$	+0.024	-0.026

carrier one can easily show that when $\delta V_L = 0$,

$$(V_B - V_0) = (1/2\beta) \ln b.$$

This follows from equations (15), (21), (22) and (38). The rest of the results hardly need comment.

It is interesting to compare differences in contact potential, etc., for units A and D. In equation (2) the Fermi energy has been included in the constant. This equation predicts that if the trap distributions on both surfaces are the same, then when the contact potentials of both surfaces are equal the values of V_D should also be equal and the difference between the V_B 's should be equal and opposite to the differences between the Fermi energies in electron volts. This equation can be written

$$\text{c.p.} = (V_B - V_0) - (V_{BO} - V_0) + V_D + V_0 + \text{const.} \quad (40)$$

In comparing units we shall use Δ 's to denote differences and these differences are always taken A-D.

For the case where $\Delta \text{c.p.} = 0$, $\Delta(V_B - V_0)$ can be read from Fig. 18 and $\Delta(V_{BO} - V_0)$ from Table II. We have then

$$\Delta(V_D + V_0 + \text{const}) = -0.05.$$

This indicates that our simple picture is not quite right. If ΔV_D were zero for this case it should follow from equation (12) that both $\Delta 2H \sinh \beta(V_B - V_{BO})$ and the difference in the const in this equation should be zero. It turns out, however, that $\Delta 2H \sinh \beta(V_B - V_{BO})$ is not zero but $+0.07$. Equation (12) can be substituted into equation (40) giving

$$\begin{aligned} \text{c.p.} = (V_B - V_0) - (V_{BO} - V_0) + 2H \sinh \beta(V_B - V_{BO}) \\ + V_0 + \text{const}_2. \end{aligned} \quad (41)$$

Where the constant now includes the constant part of V_D and is labeled const_2 to distinguish it from the constant in equation (40). We have then for $\Delta \text{c.p.} = 0$

$$\Delta(V_0 + \text{const}_2) = -0.12.$$

This difference $\Delta(V_0 + \text{const}_2)$ can be calculated three other ways, using the experiment results and theoretical values where necessary. These ways are

1) at the same time in the cycle

$$\Delta(V_0 + \text{const}_2) = -0.10$$

2) when $\delta V_L = 0$

$$\Delta(V_0 + \text{const}_2) = -0.13$$

3) when $\delta V_B = 0$

$$\Delta(V_0 + \text{const}_2) = -0.12.$$

All four results are consistent within the probable experimental accuracy and give an average result of $\Delta(V_0 + \text{const}_2) = -0.12$ instead of zero as one might expect from the simple picture. This indicates that the trap distributions are different on the two surfaces. The constant part of V_D is proportional to $-(N_a - N_b)$ and from equation (10) one would expect V_0 to decrease as N_a increases with respect to N_b , thus ΔV_0 and Δconst_2 should be of the same sign and additive so that Δconst_2 is less than 0.12 volts; assuming that ℓ_D/K_D is the order of 2×10^{-7} cm, this indicates that $\Delta(N_a - N_b)$ is the order of 3×10^{11} per cm^2 which is small compared with the probable trap density, N_a , N_b , of the order of 10^{14} per cm^2 as we shall see in the next paragraph.

Assuming that ℓ_D/K_D in equation (13a) is the order of 2×10^{-7} cm one can calculate $(n_{a0}p_{b0})^{\frac{1}{2}}$ obtaining 4.1×10^{10} and 5.5×10^{10} for samples A and D respectively. Using these values one can solve for S_t in equation (37) and obtain for the case of $V_0 = 0$ the value 5.0×10^{-17} cm^2 for the average capture cross section of the surface traps. The values of S_t , n_{a0} and p_{b0} depend on what one takes for V_0 . The relations are

$$S_t = \frac{5 \times 10^{-17}}{\cosh \beta V_0},$$

$$n_{a0} = 5 \times 10^{10} \exp [-\beta V_0],$$

$$p_{b0} = 5 \times 10^{10} \exp [\beta V_0].$$

This dependence is shown in the graph in Fig. 19. As already mentioned there are reasons for thinking that V_0 is less than zero. If one takes -0.06 volts as a reasonable value then one gets $S_t = 10^{-17}$ cm^2 , $n_{a0} = 5 \times 10^{11}/\text{cm}^2$ and $p_{a0} = 5 \times 10^9/\text{cm}^2$ respectively. One can push these calculations still further to estimate N_a , N_b , E_a and E_b . One knows that E_a and $-E_b$ must be greater than $1/kT$. Also N_a and N_b should be less than the number of germanium atoms per cm^2 of surface which is $1.4 \times 10^{15}/\text{cm}^2$. Values of N_a and N_b of the order of $1 \times 10^{14}/\text{cm}^2$ for the number of traps per cm^2 with energies E_a and $-E_b$ of the order of 0.2

e.v. measured from the midband energy are reasonable and not inconsistent with the original assumptions.

As mentioned in the experimental section, some data on $(\Delta c.p.)_L$ for samples C and E have been obtained. While no complete analysis of these results has been made, one can see that they are of the right order of magnitude. As the specific resistance of the sample decreases, the body life time τ decreases. This empirical result is to be expected.⁷ Consequently δp for the same light intensity decreases and one would expect $(\Delta c.p.)_L$ to decrease as it does. For sample E of course one could not neglect the charge in the space layer so that the theory would be more involved.

The comparison between the contact potentials of samples A, C, D and E shown in Fig. 6 can be understood in part at least. Consider first the over-all result that at the same time in the cycle the c.p. for a sand-blasted surface is less than for an etched surface, i.e., the work function for the sand-blasted surface is greater. It is known that the surface recombination increases enormously when the surface is sand-blasted. This means that either the surface trap density has increased or that the distribution has changed or both in such a way as to increase surface

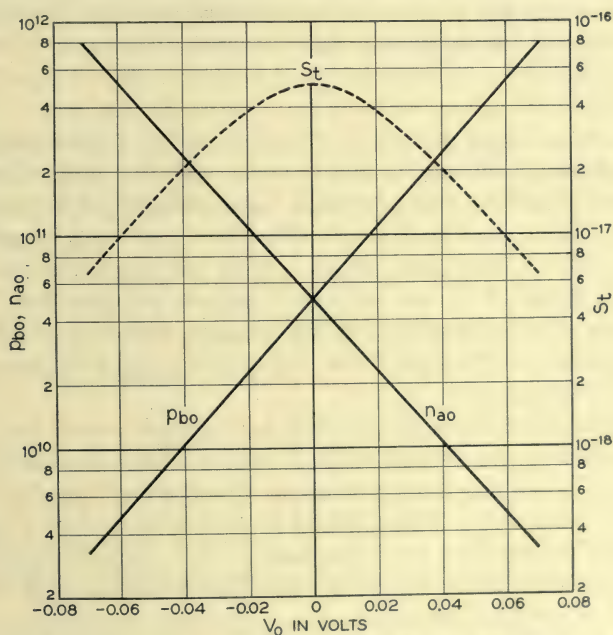


Fig. 19 — Dependence of carrier densities in surface traps, n_{a0} and P_{b0} , for $V_B = 0$, and cross section of trapping S_t on V_0 .

recombination. The results of Fig. 6 indicate that this change in trap distribution also has been in a direction to increase $(-V_B)$. This is what one would expect if N_b has increased with respect to N_a , i.e., sand-blasting increases the effective number of acceptor traps. About all that can be said about the rest of the results in Fig. 6 is that if one knew the surface trap structure in number and energy distribution for the etched surface then one could deduce from the results in Fig. 6 the distribution of traps for the sand-blasted surface over part of the energy range at least.

The theory developed here for the germanium surface may not apply at all for a silicon surface. In a previous discussion of the silicon surface¹ it was assumed that the resistance to surface trapping occurred mainly in the flow across the space charge layer rather than from the surface to the traps. This may well be the case in silicon. An investigation of the silicon surface using these same techniques should clarify this point.

It should be emphasized that the ambient used for this study and the variations in it are special in that they do not necessarily correspond to the atmosphere of ordinary room air. It is very probable that constituents in room air other than oxygen ions and water vapor are important, such as salt ions for instance. It is quite probable that v_s for such a surface exposed to room air would increase considerably with time instead of staying relatively constant as it does under the bell jar.

CONCLUSIONS

A method has been developed for studying the surface properties of Ge in a gaseous ambient at atmospheric pressure. It has been found that the Ge surface interacts with this ambient. Two atmospheric constituents that are important in this interaction, oxygen and vapors with OH radicals, have been isolated. With the controlled use of these, the surface dipole of Ge can be cycled between two extremes. Thus the dependence of other properties of the Ge surface on surface dipole, such as change of contact potential with illumination and surface recombination, can be determined.

It is evident from the results that the method is very powerful. In order to complete the present study, it was necessary to stop trying new experiments since almost all directions of variation open up new and interesting phenomena. These results on Ge are really just a beginning, and the preliminary data on silicon indicate that the same method would also be fruitful on other semiconductors. The technique is of course not limited to atmospheric pressure.

A tentative theory of the Ge surface has been developed that is sufficient to explain the experimental results on a semi-quantitative basis. Theory and the experiment together predict approximately the

number, type and distribution in energy of the surface traps. One has therefore a tentative model of the Ge surface that should be very useful in any further investigation of its properties.

ACKNOWLEDGMENTS

We wish to acknowledge the help and assistance of all our colleagues who have contributed in many ways to make this investigation a success. We wish to mention in particular E. G. Dreher and R. E.ENZ who took most of the experimental data, H. R. Moore who designed and made the electronic equipment used in making the measurements, and Conyers Herring for suggestions regarding the theory of large amplitude signals.

APPENDIX

We have assumed (Postulate IV) that traps of type *a* are in equilibrium with electrons in the conduction band and that traps of type *b* are in equilibrium with holes in the valence band. We wish to show that this is not really a separate assumption but follows as a consequence of Postulate II if the density of traps is not too high. For simplicity, we shall restrict the discussion in the appendix to the limiting case of small departures from equilibrium so that the equations are linear. The problem may then be discussed most conveniently⁷ by means of quasi-Fermi levels* or imrefs, ϕ_n and ϕ_p , for the conduction electrons and holes, respectively.

Departures $\delta\phi_n$ and $\delta\phi_p$, of the imrefs from the Fermi level are a measure of the departures of the concentrations, δn and δp , from their equilibrium values:

$$n_1 = n + \delta n = n \exp(-\beta\delta\phi_n), \quad (\text{A.1})$$

$$p_1 = p + \delta p = p \exp(\beta\delta\phi_p). \quad (\text{A.2})$$

The imrefs in the interior are then

$$\delta\phi_n = \frac{-1}{\beta} \ln \left(1 + \frac{\delta n}{n} \right), \quad (\text{A.3})$$

$$\delta\phi_p = \frac{1}{\beta} \ln \left(1 + \frac{\delta p}{p} \right). \quad (\text{A.4})$$

Correspondingly, the imrefs at the surface are defined by:

$$n_{s1} = n_s + \delta n_s = n \exp[\beta(V_B + \delta V_B - \delta\phi_{ns})], \quad (\text{A.5})$$

$$p_{s1} = p_s + \delta p_s = p \exp[-\beta(V_B + \delta V_B - \delta\phi_{ps})]. \quad (\text{A.6})$$

* Reference 2, pages 302-308.

Changes in imrefs of the traps are defined by:

$$n_t = n_i + \delta n_i = \frac{N_t}{1 + \exp [(E_t - E_F - e(V_B + \delta V_B) + e\delta\phi_t)/kT]}, \quad (\text{A.7})$$

$$p_t = p_i + \delta p_i = \frac{N_t}{1 + \exp [(E_F - E_t + e(V_B + \delta V_B) - e\delta\phi_t)/kT]}. \quad (\text{A.8})$$

In these last two equations, t may refer to either type of trap (a or b) and $\delta p_t = -\delta n_t$.

For small signals, the recombination current via a given set of surface traps may be considered as a flow produced by differences in the imrefs, $\delta\phi_n$ and $\delta\phi_p$, through four effective resistances in series, as shown in Fig. 20. Here R_{nB} is an effective resistance for flow of electrons across the barrier layer, R_{nt} for flow of electrons from the conduction band at the surface to the traps, R_{pt} for flow of holes from the filled band to the traps, and R_{pB} for flow of holes across the barrier layer. Under steady state conditions, the net flow of conduction electrons to the surface is balanced by an equal flow of holes. The recombination current may be thought of as a flow of electrons from the conduction band via the traps to the valence band.

The recombination current per unit area is

$$I = -eU = (\delta\phi_n - \delta\phi_p)/R_t, \quad (\text{A.9})$$

where R_t is the sum of the four resistances in series. Here U is the particle current and I is the corresponding electric current. If $\delta\phi_n$ and $\delta\phi_p$ are expressed in volts, R_t is in ohms/cm². We may define a recombination constant C_t by an equation corresponding to (34):

$$U = C_t(p_1 n_1 - n_i^2). \quad (\text{A.10})$$

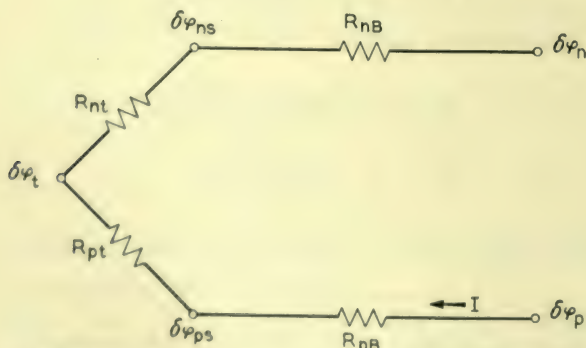


Fig. 20 — Circuit analogy of surface recombination.

The relation between C_t and R_t is obtained by use of (A.1) and (A.2) in (A.10). Since (A.9) is valid only to terms of the first order in $\delta\phi_n$ and $\delta\phi_p$, we make the corresponding approximation in (A.10) and find

$$1/R_t = e\beta n_i^2 C_t. \quad (\text{A.11})$$

If there is more than one type of trap, the net recombination resistance, R , is that of the various traps in parallel. Values obtained for R for specimens A and D from the empirical values of C are about 500 ohms.

We shall show that for a -traps, R_{pa} is much larger than the other resistances in series, and that for b -traps, R_{nb} is the dominant resistance. This implies that $\delta\phi_a = \delta\phi_n$ and $\delta\phi_b = \delta\phi_p$, or in other words that a -traps are in equilibrium with the conduction band and b -traps with the valence band.

First consider the flow across the space-charge layer. The resistances R_{nt} and R_{bt} depend on the sign of V_B . When V_B is negative, the net electron current across the space-charge layer is:

$$\begin{aligned} I &= ev_n(n_{s1} - n_1) \exp [\beta(V_B + \delta V_B)] \\ &\cong e\beta v_n n_s (\delta\phi_n - \delta\phi_{ns}), \end{aligned} \quad (\text{A.12})$$

where v_n is defined by an equation similar to (24). The second form is the linear approximation. Thus we have

$$1/R_{nB} = e\beta v_n n_s. \quad (\text{A.13})$$

If V_B is positive, n_s is replaced by n .

Correspondingly, for V_B negative, the hole current is:

$$\begin{aligned} I &= ev_p(p_1 - p_{s1}) \exp [\beta(V_B + \delta V_B)] \\ &\cong e\beta v_p p (\delta\phi_{ps} - \delta\phi_p), \end{aligned}$$

and

$$1/R_{pB} = e\beta v_p p. \quad (\text{A.14})$$

For V_B positive, p is replaced by p_s .

The maximum value of R_{nB} is obtained for the ambient which makes V_B most negative and R_{pB} is a maximum when V_B is most positive.

The current from the conduction band to the traps is calculated as in (23):

$$I = e(v_n S_{nt} n_{s1} p_{t1} - g_i n_{t1}) \quad (\text{A.15})$$

$$\cong e\beta v_n S_{nt} n_s p_t (\delta\phi_{ns} - \delta\phi_t). \quad (\text{A.16})$$

Here g_t is the rate of thermal generation of electrons from occupied traps to the conduction band. Since, for equilibrium conditions, $I = 0$, we have:

$$g_t = v_n S_{nt} n_s p_t / n_t. \quad (\text{A.17})$$

It may be verified easily that g_t is independent of E_F and V_B . The ratio,

$$n_{ct} = n_s p_t / n_t = N_c \exp [(E_t - E_c)/kT], \quad (\text{A.18})$$

is the equilibrium concentration of electrons in the conduction band when the Fermi level is at the level of the traps. The expression (A.16) is again the linear approximation. Thus

$$1/R_{nt} = e\beta v_n S_{nt} n_s p_t. \quad (\text{A.19})$$

Similarly, we have

$$1/R_{pt} = e\beta v_p S_{pt} p_s n_t. \quad (\text{A.20})$$

We shall show that the ratio

$$R_{nt}/R_{pt} = (v_n S_{pt}/v_p S_{nt})(p_s n_t/n_s p_t), \quad (\text{A.21})$$

may be expected to be small compared with unity for a -traps and large compared with unity for b -traps. First consider a -traps. If anything, $S_{pa} < S_{na}$, because holes must give up a larger energy than conduction electrons in going to a -traps. The second factor is small compared with unity if

$$p_s \ll n_{ct}, \quad (\text{A.22})$$

with n_{ct} defined by (A.18). This will be the case if the a -traps are closer to the conduction band than the top of the valence band at the surface is to the Fermi level. According to Postulate II, this should always be the case.

Similarly, for b -traps

$$R_{pt}/R_{nt} \ll 1 \quad (\text{A.23})$$

if

$$n_s \ll p_{vt}, \quad (\text{A.24})$$

where

$$p_{vt} = p_s n_t / p_t = N_v \exp [-(E_t - E_v)/kT]. \quad (\text{A.25})$$

is the concentration of holes in the valence band when the Fermi level is at the level of the traps.

The values of R_{nB} and R_{pB} relative to R_{nt} and R_{pt} are:

$$R_{nB}/R_{nt} = S_{nt}p_t \quad (\text{A.26})$$

$$R_{pB}/R_{pt} = S_{pt}n_t. \quad (\text{A.27})$$

In our experiments these are always small compared with unity, since the trapping cross-sections are of the order of 10^{-17} cm^2 and p_t and n_t are always less than N_t , which is of the order of $10^{14}/\text{cm}^2$.

Barrier resistances may be important for surfaces with a large number of surface traps. Analysis of earlier data on the change of the contact potential of silicon with light¹ was made on the assumption that the probability that an electron or hole reaching the surface be trapped is relatively large, and that consequently the barrier resistances are large compared with the trapping resistances. The present experiments on germanium throw doubt on this interpretation, but further experiments are required to clarify the situation.

REFERENCES

1. Bardeen, J., Phys. Rev., **71**, p. 717, 1947. Brattain, W. H., Phys. Rev., **72**, p. 345, 1947 and Semi-Conducting Materials Butterworths Scientific Publications Ltd., pp. 37-46, 1951.
2. Shockley, W., *Electrons and Holes in Semiconductors*. D. Van Nostrand, pp. 318-325, 1950.
3. Pearson, G. L., unpublished data. J. R. Haynes and W. Shockley, Phys. Rev. **81**, p. 835, 1951.
4. The CP-4 etch is due to R. D. Heidenreich. The formula is given in Phys. Rev., **81**, p. 838, 1951. This method of treating a Ge surface is due to C. S. Fuller.
5. C. S. Fuller suggested the use of these fumes.
6. Goucher, F. S., G. L. Pearson, M. Sparks, G. K. Teal and W. Shockley, Phys. Rev. **81**, p. 637, 1951.
7. Shockley, W. and W. T. Read, Phys. Rev., **87**, p. 835, 1952.
8. van Roosbroeck, W., Bell Sys. Tech. J., **29**, p. 560, 1950.

Frequency Economy in Mobile Radio Bands

By KENNETH BULLINGTON

(Manuscript received August 20, 1952)

The various factors affecting the usability of mobile radio channels are discussed, and estimates are obtained for the number of usable channels per megacycle for several present and proposed methods of operation. The lack of radio-frequency selectivity is the principal barrier to maximum frequency economy, but this difficulty can be avoided by sufficient geographical and operational coordination.

The increasing demand for all types of radio services emphasizes the need for efficient use of the radio frequency spectrum. In mobile radio operation the number of usable channels that can be obtained in the VHF and UHF mobile bands depends not only on the width of the individual channels, but also on how and where each channel is to be used. Activity on the same frequency at neighboring locations, and on neighboring frequencies at the same location both affect the usefulness of a channel. Halving the channel spacing doubles the number of potential assignments, but it does not double, and in some cases it does not appreciably increase the number of usable channels.

The usefulness of a single isolated channel is determined by the intensity of its signal above the noise level. Because of the very wide variation in received signal strength caused by distance, terrain, building shadowing, etc., the coverage area of a channel can be discussed only in statistical terms. There are likely to be islands of poor signal-to-noise ratio even close to the transmitter, and the coverage gradually fades out into more spotty conditions at greater distance.

If the same frequency is used at a neighboring location, the familiar problem of co-channel interference arises. There will now be locations where the desired signal is above noise, but the undesired signal is still stronger. Thus, the coverage area of a channel is reduced by the existence of the co-channel transmitter; again, it is possible to discuss this reduction only in statistical terms.

When two channels are being operated on different frequencies in the same general area, the coverage area of each is limited by signal-to-

noise considerations. In addition, each channel may affect the other because of spurious radiation from transmitters, insufficient receiver selectivity, receiver oscillator radiation, etc. The recent trend toward receivers with greatly improved IF selectivity is worthwhile, but even infinite IF selectivity cannot solve many of the present interference problems.

When three or more channels are operating in the same general area, another type of interference occurs because of intermodulation in transmitters or receivers. If it were technically feasible to build into the equipment sufficient radio frequency selectivity to separate the working channels, this interference could be removed. In fact, this is not feasible, and it is necessary to consider possible modulation products from channels falling within a frequency band several percent wide. The number of possible interference conditions that result from intermodulation (third order) rises from 9 for 3 working channels to 50 for 5 channels, to 450 for 10 channels, and to 495,000 for 100 working channels. Some of these interference combinations overlap and fall on the same channel; but even considering all possible duplication, intermodulation interference rapidly becomes controlling as the number of closely spaced channels working in the same area is increased.

It is not technically feasible to achieve enough radio frequency selectivity to permit unrestricted and uncoordinated use of many channels in a given area, unless the channels are, on the average, separated by about 1 per cent of the operating frequency. For any kind of efficiency of frequency utilization, it is necessary to have some coordination in the location of fixed transmitters and in the use of channels. The maximum efficiency of utilization requires the maximum coordination.

The technical factors that determine channel width, channel spacing, and the number of usable channels are described and tabulated below. The first section discusses the principal factors that affect the usefulness of channels equipped with transmitters and receivers with perfect filtering. This is followed by a consideration of the limitations imposed by insufficient total filtering and by insufficient radio frequency filtering. The next section shows the reduced requirements that are possible by coordination between systems. Finally, the quantitative data are used to illustrate the capabilities and efficiencies of various present and proposed methods of mobile system operation.

CHANNELS WITH PERFECT FILTERING

It has been found by experiment that the radio path loss between antennas in a mobile radio system can be ascribed to three principal fac-

tors: distance, shadowing and standing wave patterns. The variation with distance from the base station follows the theoretical free space loss up to 500 feet or more, as long as the points are within line of sight. Typical values of the free-space loss are shown in Table I. Beyond about one-half mile the median path loss over plane earth increases about 12 db each time the distance is doubled out to distances of 20–30 miles.^{1, 2}

In addition to the increase in path loss with distance, which is accounted for reasonably well by the theory of radio propagation over plane earth, bold features of geography such as mountains and large buildings cause shadow losses that result in irregular coverage patterns. For example the median loss at street level for random locations in New York City is about 25 db greater than the plane earth values computed for the distance and antenna heights involved; the corresponding 10 per cent and 90 per cent losses are about 15 and 35 db respectively.³

Superimposed on the above effects which vary relatively slowly with location are standing wave patterns whose effect on path loss can change substantially within a foot or so. The standing waves are the result of random additions of multiple reflections from nearby buildings or terrain, and the variation in path loss follows the Rayleigh distribution for small changes in distance in urban areas. In other words, there is no theoretical limit on the deviation from the median but in 1 per cent of the possible locations the signal is likely to be more than 8 db above the median value and in 99 per cent of the possible locations the signal level is not expected to be more than 18 db below the median value.

The motion of the mobile unit through the standing wave patterns causes signal fluctuations or flutter in the received signal. The flutter

TABLE I — FREE SPACE LOSS BETWEEN DIPOLES

Separation Between Transmitting and Receiving Antennas	Free-Space Loss	
	150 mc	450 mc
5 ft.	16 db	26 db
50 ft.	36	46
500 ft.	56	66
$\frac{1}{2}$ mile	70	80

¹ Young, W. R., Jr., Comparison of Mobile Radio Transmission at 150, 450, 900 and 3700 Mc. Bell Sys. Tech. J., 30, pp. 1068–1085, Nov., 1952.

² Aikens, A. J., and L. Y. Lacy, A Test of 450-Megacycle. Urban Transmission to a Mobile Receiver. I.R.E., Proc., pp. 1317–1319, Nov., 1950.

³ Bullington, K., Radio Propagation Variations at VHF and UHF. I.R.E., Proc., pp. 27–32, Jan., 1950.

rate at 150 mc may be as much as 15 cycles per second for a speed of 30 mph and increases as either the radio frequency or the speed of the mobile unit is increased. The fast acting gain control needed to minimize the flutter effects is obtained automatically with frequency modulation but is more difficult to obtain with amplitude modulation. This factor is one of the principal advantages of the use of FM instead of AM for mobile radio systems.

The co-channel interference to be expected between stations having equal transmitter powers depends on the path loss statistics for both the desired and undesired signals. At the edge of the desired coverage area there must be a high probability that the desired signal will be strong enough to be useful and only a small probability that the undesired signal will be strong enough to be troublesome. The geographical separation needed between co-channel stations varies from about four to six times the desired coverage radius when FM is used and from six to eight times when AM is used.⁴ If the needs for mobile channels were uniformly distributed geographically only a small part of the potential channel assignments would ever be used in a given area. However, the needs for mobile channels are usually concentrated in areas of high population density so that a large percentage of the channel assignments may be needed in the same area.

The above estimates on co-channel spacing depend somewhat on the antenna heights and the type of terrain, and assume that the same frequency is used in both directions of transmission. When the two-frequency method is used with adequate separation between the transmitting and receiving frequencies, the co-channel spacings can be reduced to about three to five times the coverage radius for FM and to about four to six times for AM. This reduction of approximately 30 per cent is possible because the most troublesome interfering path in the single frequency method (from base transmitter to base receiver) can be eliminated in the two-frequency method by sufficient selectivity.

The principal reason for using the single frequency method is to provide communication between two mobile units when they are relatively near each other but are beyond the range of the base station. When transmission of all messages through the base station is desirable, or at least not objectionable, the two-frequency method is preferable. It is shown in a later section that close geographical and operational coordination is needed to achieve maximum efficiency in the use of frequency space and that this coordination can be obtained only with the two-frequency method.

⁴ See reference in Footnote 3.

The bandwidth needed to pass the desired signal depends on the frequency stability that can be maintained as well as on the type of modulation. The allowance for frequency drift includes the variations in both transmitters and receivers. The importance of these figures is indicated in Table II which shows the tolerances needed for frequency instability. For example, with an overall frequency stability of ± 0.002 per cent the channel width at 450 mc needs to be 18 kc wider than the minimum bandwidth required to pass the modulated signals.

The use of frequency modulation has several important advantages that cannot be readily obtained with AM. The instantaneous gain control and the closer co-channel spacings have already been mentioned. In addition, for the same radiated power, FM with a frequency swing greater than about ± 3 kc has the well known advantage of providing a higher output signal-to-noise ratio throughout most of the coverage area than is possible with double sideband amplitude modulation; this FM advantage is substantially reduced when the IF bandwidth is large compared with the bandwidth required to pass the desired sidebands.

The bandwidth required for frequency modulation of a 3 kc voice band must be at least ± 3 kc. For reasonable FM signal-to-noise advantage, particularly in the presence of impulse noise, the frequency swing should be at least ± 5 kc which requires a bandwidth of ± 8 kc for good quality. The corresponding bandwidth for amplitude modulation is ± 3 kc; the use of single sideband AM transmission does not seem feasible, at least not for single channel operation.

LIMITATIONS IMPOSED BY INSUFFICIENT (TOTAL) FILTERING

The frequency separation between carrier frequencies must be greater than the bandwidth required to pass the desired signal because additional frequency space or guard bands are needed to build up receiver selectivity against undesired signals and to avoid the extra band radiation from transmitters. The power of a 100 watt transmitter is about

TABLE II — TOLERANCE NEEDED FOR OVERALL FREQUENCY DRIFT

Frequency Stability	Allowance for Frequency Drift	
	150 mc Band	450 mc Band
$\pm 0.001\%$	± 1.5 kc	± 4.5 kc
± 0.002	± 3	± 9
± 0.005	± 7.5	± 22.5

TABLE III — REQUIRED SUPPRESSION VERSUS DISTANCE
BETWEEN ANTENNAS

Distance Between Transmitting and Receiving Antennas	Total Selectivity or Filtering Required	
	150 mc	450 mc
0 ft.	160 db	160 db
50 ft.	124	114
500 ft.	104	94
$\frac{1}{2}$ mile	90	80

160 db greater than the minimum signal that is useful in the receiver (140 db below one watt), so ideally no appreciable interference would result if the overall selectivity of the receiver and the suppression of extra band radiation in the transmitter could be in excess of 160 db. This amount of isolation is difficult to obtain by filtering. The interaction between transmitter and receiver of the same system is frequently avoided by the use of "push-to-talk" operation, but the potential interference between different systems requires the full 160 db (based on 100 watt transmitters). Fortunately, a substantial part of the desired isolation can be obtained by modest geographical separation. The net requirements for either receiver selectivity or transmitter filtering are less than 160 db by the losses shown in Table I and are summarized in Table III.

Receiver selectivities of 90–100 db or more are feasible except on nearby channels and possibly on certain image channels. Typical values of the guard bands that are required between the edge of the desired pass band and the frequency at which the desired attenuation to interfering signals can be obtained are estimated in Table IV.

Even if the guard band, shown in Table IV, required to provide adequate selectivity in the receiver could be reduced to zero by providing infinitely steep sides on the IF selectivity curve, there would still remain the guard band needed to avoid the extra band radiation from the transmitter. The amount of suppression of extra band radiation needed

TABLE IV — GUARD BAND VERSUS IF SELECTIVITY

Desired IF Selectivity	Required Guard Band
40 db	12 kc
60	15
80	20
100	25
120	30

for unrestricted operation is equal to the required receiver selectivity given in Table III and can be translated into frequency space in the following manner.

Both AM and FM transmitters radiate some noise and distortion products outside of the ideal modulation bandwidth. In addition, some of the sideband energy in FM falls outside the desired modulation bandwidth. The magnitude of the undesired FM sideband radiation is higher than the noise immediately outside of the desired band, but it decreases more rapidly with the result that the noise is usually controlling in the region where the extra band radiation is more than 60 to 70 db down.

The guard bands that are required between the edge of the desired transmitted band and the frequency at which the necessary suppression of extra band radiation can be obtained are estimated in Table V. These values depend on the width of the voice band and are relatively independent of the radio frequency since r.f. selectivity is not possible.

Measurements on present day transmitters correspond to the above estimates for values of suppression less than 80 db, but a frequency separation of nearly one megacycle or more is needed for suppressions of 100 and 120 db. This limitation is not expected to be inherent so more optimistic estimates are indicated in Table V. If the present characteristics cannot be improved, that is, if suppressions greater than about 80 db cannot be obtained, Table III indicates that some interference may be expected within about one-half mile of an unwanted transmitter.

A comparison of the information given in Tables III, IV and V indicates that the guard bands required for unrestricted operation are approximately 100, 50 and 25 kc for minimum separations between transmitter and receiver of 50 feet, 500 feet and one-half mile, respectively. These values together with the bandwidth needed for modulation and for frequency instability determine the frequency separation required between channels operating in the same area and are summarized in Table VI.

TABLE V — GUARD BANDS REQUIRED TO AVOID EXTRA BAND RADIATION

Suppression of Extra Band Radiation	Guard Bands Required	
	AM	FM
40 db	3 kc	9 kc
60	10	15
80	25	25
100	50	50
120	100	100

TABLE VI — CHANNEL SPACING REQUIRED FOR UNRESTRICTED
OPERATION OF TWO FM CHANNELS IN SAME AREA
VERSUS ANTENNA SEPARATION

Minimum Separation Between Transmitting and Receiving Antenna	Channel Spacing, Neglecting Intermodulation			
	150 mc		450 mc	
	$\pm 0.002\%$	$\pm 0.005\%$	$\pm 0.002\%$	$\pm 0.005\%$
50 ft.	112-122 kc	121-131 kc	124-134 kc	151-161 kc
500 ft.	62- 72	71- 81	74- 84	101-111
$\frac{1}{2}$ mile	37- 47	46- 56	49- 59	76- 86

The above table shows that if interference of the types so far considered is to be kept below the minimum usable signal at all distances greater than about 500 feet from undesired transmitters, the channel spacing needs to be at least 62 to 75 kc in the 150 mc band and 74 to 105 kc in the 450 mc band. The channel spacings for AM are equal to the minimum shown above, while the higher figure is for FM with ± 5 kc swing (a modulation bandwidth of ± 8 kc).

Since the above channel spacings are considerably greater than the necessary IF bandwidth, it should be possible to use intermediate channels in adjacent non-overlapping areas. This geographical limitation does not appreciably decrease the overall efficiency in the use of frequency space as long as the needs for mobile channels are more or less uniformly distributed within a large region, but it becomes important where a large percentage of the available channels are needed in the same metropolitan area.

In a later section it is shown that channel spacings less than the values given in Table VI are feasible in the same area providing sufficient co-ordination is achieved in both geographical spacings and operating methods.

The estimated channel spacings shown in Table VI do not take into account the effect of intermodulation interference which is discussed in the following section. Intermodulation interference may limit the number of usable one-way channels to only 1 or 2 per megacycle instead of the above 6 to 20 per megacycle, unless further restrictions are placed on the selection of frequencies and on the method of operation.

LIMITATIONS IMPOSED BY INSUFFICIENT RF FILTERING

When a strong unwanted signal on a frequency within the RF bandwidth is present at the input to a receiver, overloading occurs and the receiver

TABLE VII — REQUIRED RF RECEIVER SELECTIVITY VERSUS ANTENNA SEPARATION

Minimum Separation Between Transmitter and Receiver	RF Selectivity	
	150 mc	450 mc
0 ft. (common antenna)	95 db	95 db
50 ft.	59	49
500 ft.	39	29
$\frac{1}{2}$ mile	25	15

is said to be desensitized. When two or more strong unwanted signals are present desensitization also occurs, but in addition, extraneous frequencies are generated by intermodulation in the receiver itself. As the levels of the unwanted signals become greater than about 75 db below one watt (1 or 2 millivolts across a typical receiver) the intensity of the modulation products rises rapidly above the set noise. The resulting interference can be 60 db or more above set noise and the number of the modulation products increases by at least the cube power of the number of operating channels.

Ideally, the intermodulation interference in the receiver caused by 100-watt transmitters (20 db above one watt) can be eliminated by $20 + 75 = 95$ db RF selectivity even when the receiver and the unwanted transmitters are connected to the same antenna. In practice, the effect of geographical separation assuming the free space loss given in Table I reduces the RF selectivity requirement to the values given in Table VII.

The RF selectivity requirements given in Table VII cannot be obtained on nearby channels. The approximate RF bandwidths associated with various amounts of RF selectivity in mobile receivers is shown in Table VIII. For example, in mobile receivers it seems feasible to provide 40 db of RF selectivity at frequencies removed from the desired channel by about 3 mc in the 150-mc band and by about 10 mc in the 450-mc band. At fixed stations the RF bandwidth required for a given selec-

TABLE VIII — FREQUENCY SPACING FROM MIDBAND VERSUS RF SELECTIVITY

Desired RF Selectivity	Frequency Spacing from Midband	
	150 mc	450 mc
20 db	± 1.5 mc	± 5 mc
40	± 3	± 10
60	± 6	± 20

TABLE IX — SIGNIFICANT RF BAND VERSUS ANTENNA SPACING

Minimum Separation Between Receiver and Unwanted Transmitters	RF Band	
	150 mc	450 mc
50 ft	± 6 mc	± 14 mc
500 ft.	± 3	± 7
$\frac{1}{2}$ mile	± 2	± 4

tivity can be reduced to one-third and possibly to one-fourth of the above values by the use of bulky and expensive filters.

The critical frequency band that needs to be considered in determining the usefulness of any given channel can be obtained by combining the information given in the two preceding tables with the results shown in Table IX. For example, if it be desired to work mobile receivers unrestricted to within 500 feet of two or more unwanted transmitters, all frequency assignments within ± 3 mc in the 150-mc band (or within ± 7 mc in the 450-mc band) must be carefully chosen if intermodulation interference is to be avoided.

When the ± 3 -mc band is divided into 100 potential channel assignments of 60 kc each and when the channels assigned to a given area are chosen at random, 7 channels working 50 per cent of the time (or 37 channels working 10 per cent of the time) will, on the average, cause third order intermodulation interference about 10 per cent of the time on each channel within the band. The interference is expected to be above the minimum usable signal level in all receivers located less than about a mile from the unwanted transmitters. Even if the operating frequencies are selected carefully instead of at random, no more than 11 channels out of 100 can be found that are free of third order intermodulation when used simultaneously in the same general area. These results are discussed more completely in a companion paper.⁵ When the number of potential channel assignments is greater or less than 100, the corresponding number of usable channels limited by third order modulation alone is shown in Table X. The numbers of usable channels shown above are further reduced when fifth and higher order intermodulation products are considered.

A reduction in the nominal channel spacing from 60 kc to 20 kc means a three-fold increase in the potential channel assignments, but Table X shows that the number of usable channels increases much more slowly.

⁵ Babcock, W. C., Intermodulation Interference in Radio Systems. Page 63 of this issue.

TABLE X — NUMBER OF USABLE CHANNELS VERSUS NUMBER OF POTENTIAL CHANNELS

No. of Potential Channel Assignments in RF Band Shown in Table IX	No. of Usable Channels			
	Careful Selection No Interference	Random Selection 10% Chance of Interference		
		% of Time Transmitter Is On		
		50%	25%	10%
20	7	5	10	25
50	9	6	12	30
100	11	7	15	37
200	12	9	18	45
500	14	12	24	60

Thus far, only the intermodulation interference generated in the receivers has been considered. Intermodulation also occurs at the same frequencies in the transmitters, but it usually can be made less important than the corresponding interference in the receivers. Ideally, the intermodulation products generated in the transmitters should not be stronger than 140 db below one watt (about 1 microvolt at the input to the receiver) which requires about 75 db RF filtering in each transmitter output. This ideal requirement is based on 100 watt transmitters with both the transmitters and receiver working on the same antenna. In practice, the RF filtering requirement is less than 75 db because of physical separation between transmitters and receivers, and typical values based on free space transmission are shown in Table XI.

A comparison of the filter requirements on 100 watt transmitters with the corresponding receiver selectivity requirements given in Table VII shows that the receiver requirements are greater as long as the effective

TABLE XI — RF TRANSMITTER FILTERING VERSUS ANTENNA SEPARATION

Distance Between Receiver and Unwanted Transmitters	RF Filtering Needed in Each Transmitter in db							
	150 mc				450 mc			
	Distance Between Transmitters				Distance Between Transmitters			
	0 ft	10 ft	50 ft	500 ft	0 ft	10 ft	50 ft	500 ft
0 ft.*	75	—	—	—	75	—	—	—
50 ft.	57	46	39	—	52	36	29	—
500 ft.	47	36	29	19	42	26	19	9
$\frac{1}{2}$ mile	40	29	22	12	35	19	12	2

* Common antenna.

separation between transmitters is greater than about 50 feet. For example, with a 500-foot separation between the transmitting and receiving antennas, Table VII shows that the 150 mc requirement on r.f. selectivity is 39 db. The bandwidth between the 39 db points on the receiver selectivity characteristic determines the number of potential channel assignments to be used in Table X.

INCREASED EFFICIENCY OBTAINED BY COORDINATION

The preceding selectivity and filtering requirements are severe and in some cases virtually unattainable except at considerable sacrifice in frequency space. The principal reason for these exacting requirements is that the assumed unrestricted and independent operation results in large differences in field intensities among closely spaced frequencies. In order to pick out the weak signals from among the strong, sufficient selectivity must be provided to suppress the potential interference to below the minimum usable signal.

An alternative is to reduce the level differences and hence the filtering requirements by geographical and operational coordination. This means that the level of the potential interference can be permitted to be many db above set noise as long as it is always at least 10-20 db below the desired signal at all possible locations. By proper coordination the troublesome RF filtering problems can be eliminated within the coordinated system and the remaining IF selectivity problems can be minimized.

The first step is to use the two frequency method of operation with adequate separation between the frequencies used for the opposite directions of transmission. In this way substantial RF filtering can be obtained to eliminate the interference between one or more base transmitters and a base receiver. This type of interference is particularly troublesome between single frequency systems because of the relatively high base transmitter power and because the high antennas at both locations reduce the radio path loss to a minimum. The corresponding possible interference between transmitters and receivers on different mobile units is also reduced by the two frequency method but interference between mobile units is much less important because of the lower power and much lower antenna heights.

The potential interference between base transmitters and mobile receivers caused by insufficient total filtering can be reduced by locating all base transmitters at or near a common point so the level differences between the desired and undesired signals will never be excessive. When all transmitters radiate from a common antenna, a selec-

tivity or filtering requirement of about 40 db (instead of the values shown in Table III) is sufficient for a reasonable signal to interference ratio plus an allowance for differential path losses resulting from standing wave effects.

The RF selectivity or intermodulation problem in the mobile receiver can be eliminated by reducing the power level at the first converter to about 75 db below one watt. This can be done by providing a simple automatic gain control in the RF stage of the mobile receiver. In regions where the desired and undesired signals are weak the receiver has full sensitivity, while at locations near the transmitters both the desired and undesired signals are reduced in level before reaching the first converter. The result is that the intermodulation products generated in the receiver are reduced about 3 db for every db that the desired signal is lowered and the distortion becomes negligible before the output signal-to-noise ratio is reduced appreciably. In order that the a.g.c. circuit can be fully effective it is necessary that the transmitters be grouped together and that the desired carrier be transmitted to control the gain of the receiver.

Grouping the base transmitters at or near a common point together with the associated measures of transmitting the carriers and using a.g.c. greatly reduces the requirements on the mobile receiver, but these measures complicate the design of the base transmitter. The intermodulation products generated in the closely associated transmitters result in potential interference both within and outside of the desired transmitting band. The intermodulation that falls on the mobile receiver frequencies needs to be suppressed by at least 25 db below the carrier on any channel to prevent mutual interference within the coordinated system. The intermodulation that appears as extra band radiation outside the frequency range of the coordinated system must be suppressed by RF filters. The guard band needed to prevent mutual interference between the coordinated system and its neighbors is small compared with the frequency space that is saved by the close spacing of the channels within the coordinated system.

In the direction of transmission from the mobile transmitters to the base receivers, the above coordinating methods cannot be used but equally effective ones are available. The RF selectivity requirements shown in Table VII can be reduced 20 db by using 20 db less power in the mobile transmitter than in the base transmitter. This measure is somewhat analogous to the use of a.g.c. in the opposite direction of transmission; a further step would be automatic control of the radiated power but this complication does not appear to be necessary.

In order to regain the full coverage area, multiple base receivers at

different locations are needed and this use of space diversity techniques provides an opportunity to pick the receiver having the best signal-to-noise ratio. Moreover, the low power in the mobile transmitter together with the better RF filters that are possible in fixed locations reduces the critical bandwidth within which intermodulation interference can arise to about ± 0.4 mc at 150 mc and to about ± 0.6 mc in the 450 mc range. In these bandwidths approximately 20–25 channels can be obtained which with random location of the mobile units would be divided more or less uniformly among five or more base receiving stations. Since no more than 4 or 5 channels would be operating within the critical RF bandwidth at any one receiving location, the possibility of intermodulation interference is almost negligible. Finally, an off-channel squelch circuit is provided which disables the base receiver at a location where serious adjacent channel interference is most likely to occur and forces the choice of another base receiver in a different location. Another effect of the off-channel squelch circuit is that it keeps the base receiver quiet during idle times, and in this respect it is analogous to the advantage gained in the mobile receiver by continuous transmission of the desired carrier at the base transmitter.

Most of the above coordinating methods tend to emphasize and to increase the characteristic differences between the two directions of transmission. The net effects are that greater frequency economy is obtained and that the electrical requirements are reduced on the mobile equipment where size, weight and power are critical and where cost savings are important because of the large number involved. An increase in complexity occurs at the multi-channel base station but this seems economically justified because the cost can be divided among many working channels.

When the above methods of coordination are fully utilized, the RF requirements are eliminated in the mobile equipment and can be met in the base station equipment. In addition, the IF selectivity requirement on nearby channels is reduced to about 40 db in the mobile receiver and to about 60 db in the base receiver. The extra band radiation requirement on nearby channels is reduced to about 25–40 db in the base transmitter, depending on whether one or more than one antenna is used; and to about 60 db in the mobile transmitter.

These requirements coupled with the data given in Tables II, IV and V lead to the frequency separation between coordinated channels operating in the same area as given in Table XII. The channel spacings are shown for AM and for FM with a frequency swing of ± 5 kc (which requires a bandwidth of ± 8 kc for good quality).

The spacings shown in Table XII assume that each channel is trans-

TABLE XII—CHANNEL SPACING VERSUS SYSTEM STABILITY—
COORDINATED SYSTEMS IN SAME AREAS

Stability	Channel Spacing							
	150 mc				450 mc			
	Mobile Receiver		Base Receiver		Mobile Receiver		Bass Receiver	
	AM	FM	AM	FM	AM	FM	AM	FM
$\pm 0.001\%$	21	31	25	35	27	37	31	41
$\pm 0.002\%$	24	34	28	38	36	46	40	50
$\pm 0.005\%$	33	43	37	47	63	73	67	77

mitted on an individual carrier. Single-channel operation seems to be the only practical arrangement for transmission from the mobile transmitter to the base receiver. In the other direction of transmission, from base transmitter to mobile receiver, the question naturally arises whether additional frequency economy could be achieved by multichannel methods. In this case individual carrier operation is also indicated for transmission and economic reasons. The multiple echoes that exist at street level in urban areas limit the number of usable channels that can be transmitted on a single carrier.⁶ While the exact number is somewhat indefinite, it appears to be less than about 20 and perhaps less than 10 channels. In addition the selectivity and linearity requirements on multi-channel receivers (even for two channels) are much more severe than for single channel equipment. From these considerations it appears that the use of more expensive receivers and channel separation equipment in each mobile unit is not economically feasible.

FREQUENCY ECONOMY IN PRESENT AND PROPOSED MOBILE SYSTEMS

The technical factors given above provide a basis for estimating the number of usable mobile channels that can be obtained in a given bandwidth. This bandwidth must be sufficiently large to be isolated by RF filtering if the results are to be well defined.

The following examples assume two different geographical distributions: (1) the number of usable channels with overlapping coverage areas that can be obtained within a city or metropolitan area, and (2) the number of usable channels that can be obtained when the channels are distributed more or less uniformly over a state or other large area. The examples are based on the use of frequency modulation with a

⁶ Young, W. R., Jr., and L. Y. Lacy, Echoes in Transmission at 450 Megacycles from Land-to-Car Radio Units. I.R.E., Proc., pp. 255-258, March, 1950.

modulation bandwidth of ± 8 kc and a frequency stability of $\pm .002$ per cent; with these assumptions, the IF passband should be at least 22 kc in the 150 mc band and 34 kc in the 450 mc band. Narrower bandwidths could be used but this would result in a substantial sacrifice in coverage under impulse noise conditions.

Five cases are considered:

(1) *Single Frequency Semi-Coordinated*—In this case, substantially no interference is expected from third order modulation problems, which are avoided by careful selection of operating frequencies, but higher order modulation products may be important. Base station locations are unrelated geographically to other systems in same general area, except that a minimum spacing of 500 feet between receiver and interfering transmitter is assumed.

(2) *Single Frequency with Interference*—In this case, the choice of frequencies is unrestricted, but a 10 per cent chance of third order intermodulation interference is accepted within 500 feet of unwanted transmitters, when transmitters are in operation 25 per cent of time.

(3) *Two Frequency Semi-coordinated*—This is the same as (1), except with two-frequency operation.

(4) *Two Frequency with Interference*—Same as (2) except with two frequency operation.

(5) *Fully Coordinated Broad-band*—This case assumes: (a) two frequency operation with the land transmitters coordinated in location, power, antenna height and emission of protective carriers; (b) low power mobile transmitters; (c) multiple land receivers; (d) no interference from third or higher order intermodulation; and (e) guard bands to protect mobile and neighboring services from mutual interference.

The number of usable channels that can be obtained in the same area is estimated in Table XIII for frequencies near 150 mc.

The minimum channel spacing shown in the first column of Table XIII is calculated as follows: in cases (1), (2), (3) and (4), the extra band radiation from the base transmitter is controlling. As shown in Tables III and V, to avoid interference for distances greater than 500 feet from the interfering transmitter requires a guard band of about 50 kc. This is added to the 22 kc required IF pass-band of which ± 8 kc is allowed for the FM signal, and ± 3 kc for 0.002 per cent system instability. In (5), the adjacent channel receiver selectivity is controlling: Table IV shows the required 60 db can be obtained in 15 kc, which added to the required 22 kc IF band gives approximately 40 kc.

It will be noted from Table VII that the assumption of a separation of 500 feet between the receiver and the interfering transmitter requires

TABLE XIII — USABLE CHANNELS IN CITY AT 150 MC

Method of Operation	Minimum (Not Average) Channel Spacing in Same Area	Number of Usable Channels in 6 mc
(1) Single frequency semi-coordinated.....	75 kc	10
(2) Single frequency with interference.....	75	14
(3) Two frequency semi-coordinated.....	75	5
(4) Two frequency with interference.....	75	7
(5) Fully coordinated broad-band*.....	40	45

* Includes three guard bands of 0.8 mc each to protect mobile and neighboring services from mutual interference.

about 40 db RF selectivity, and from Table VIII that the 40 db selectivity requires that all frequencies within ± 3 mc need to be considered. With 75 kc channel spacing, there are 80 potential assignments in 6 mc. Table X indicates that 10 one-way channels can be found that are free of mutual third order intermodulation interference. If the available bandwidth were 12 mc the number of interference-free channels would be doubled.

By the same process from Table X, we derive the number of usable channels shown for case (2).

For cases (3) and (4), the methods are the same, but the number of usable channels is reduced to one-half that shown for the single frequency cases.

In the fully coordinated broad-band system (case 5) a usable one-way channel can be obtained every 40 kc. However, three guard bands totaling 2.4 mc are provided to protect both the mobile and neighboring systems from mutual interference. If the available bandwidth were 12 mc the number of interference-free channels would be increased from 45 to 120 since no additional guard bands would be required.

The comparison between various methods of operation given in Table XIII applies to 150-mc channels operating in the same city. When the channels are distributed more or less uniformly over a large area, the number of usable channels is increased by several factors. The separation between carrier frequencies in non-overlapping areas needs to be only slightly greater than the IF pass-band of the receiver, say, 30 kc at 150 mc. The guard bands needed in one location can be used in other areas at geographical separations less than co-channel spacings. Finally the required geographical separation between co-channel stations is less for the two frequency method than for the single frequency method and is less for FM than it would be for AM.

An estimate of the maximum number of usable channels within a large

TABLE XIV — USABLE CHANNELS IN STATE OR LARGE AREA
AT 150 MC

Method of Operation	Minimum Channel Spacing*	Number of Usable Channels in 6 mc
(1) Single frequency semi-coordinated.....	25 kc	108
(2) Single frequency with interference.....	25	171
(3) Two frequency semi-coordinated.....	25	108
(4) Two frequency with interference.....	25	171
(5) Fully coordinated broad-band.....	25	240

* Assumes adjacent channels are not assigned in same area.

area can be obtained by considering an area whose radius is about six times the coverage radius of the individual transmitter. A larger area is unnecessary because single frequency FM channel assignments can be repeated at this distance, while a smaller area would tend to approach the common area concept used above. The large area can be divided into 9 subareas, each of which can be treated in the manner used in Table XIII. The results are shown in Table XIV, which again assumes an FM modulation bandwidth of ± 8 kc and ± 0.002 per cent overall system frequency stability.

The entries in Table XIV are calculated as follows: Once again, the smallest band to be considered is limited by the RF selectivity in mobile receivers to 6 mc; with 25 kc as the minimum channel spacing, there are $6000/25 = 240$ potential assignments. From Table X, only 12 can be found to be free of third order intermodulation. With 12 channels in each of 9 subareas, there is a grand total of 108 channels usable in the state or large area. With more frequency space, the usable number is increased in proportion.

By the same process, from Table X we derive the number shown for case (2).

In the two frequency cases, the co-channel separation can be made smaller than in the single frequency cases, since the most troublesome case of interference (that between base transmitters and base receivers) is eased by RF selectivity. Thus, the co-channel separation needs to be only about 0.7 that for single frequency operation, which means that there are now effectively 18 instead of 9 subareas. It follows that the grand total of usable channels is the same in cases (1) and (3) and cases (2) and (4).

In considering case (5), we note from Table XIII that 40 kc is the minimum channel spacing usable in a single subarea. However, the largest grand total of channels is found by using 50 kc spacing in the subareas, and assigning the adjacent 25 kc channels to other subareas.

TABLE XV — USABLE CHANNELS IN CITY AT 450 MC

Method of Operation	Minimum (Not Average) Channel Spacing	Number of Usable Channels in 14 mc
(1) Single frequency semi-coordinated.....	85 kc	12
(2) Single frequency with interference.....	85	17
(3) Two frequency semi-coordinated.....	85	6
(4) Two frequency with interference.....	85	9
(5) Fully coordinated broad-band*.....	50	68

* Includes three guard bands of 2.4 mc each to protect mobile and neighboring services from mutual interference.

Similarly, the guard bands of one subarea can be used for channels elsewhere so all of the available 240 channels can be used.

The examples given in Tables XIII and XIV represent the two extreme conditions and the practical situation lies in between the two.

By similar reasoning it is possible to estimate the number of usable channels that can be obtained at frequencies around 450 mc. The number of usable channels shown in Table XV is for overlapping coverage areas in a city or metropolitan area and the estimates given in Table XVI are based on a uniform distribution over a state or other large section of the country.

Again, FM modulation with a bandwidth of ± 8 kc and a system frequency stability of 0.002 per cent are assumed.

For a bandwidth of 28 mc instead of 14 mc the number of usable channels is doubled for the first four cases and is increased from 68 to 208 for the fifth case. The corresponding estimates for bandwidths less than 14 mc are indefinite because of insufficient r.f. selectivity.

CONCLUSIONS

The principal conclusions that result from Tables XIII, XIV, XV and XVI, and from the preceding discussion can be summarized as fol-

TABLE XVI — USABLE CHANNELS IN STATE OR LARGE AREA
AT 450 MC

Method of Operation	Minimum Channel Spacing*	Number of Usable Channels in 14 mc
(1) Single frequency semi-coordinated.....	35 kc	117
(2) Single frequency with interference.....	35	198
(3) Two frequency semi-coordinated.....	35	117
(4) Two frequency with interference.....	35	198
(5) Fully coordinated broad-band.....	35	400

* Assumes adjacent channels are not assigned in same area.

lows:

1. A fully coordinated system requires a band of several megacycles that can be treated as a unit, but it offers substantial overall frequency economy and freedom from interference that can be obtained in no other way. This is particularly true in large metropolitan areas where the demand is greatest. With the same equipment and the same standards of quality and reliability, coordinated channels can always be spaced much closer in frequency than uncoordinated systems.

2. The advantages of coordination increase rapidly as the number of channels per unit area is increased. However, in areas where only three or four channels are required, the advantages of complete coordination are sufficiently small that only the semi-coordination of careful frequency allocation is required to preserve overall frequency economy.

3. For maximum economy, where full coordination is not used, the channels should be assigned as in FM and TV broadcasting first to areas and then to users within areas. The allocation of a block of channels to a particular service with a minimum of operational and geographical restriction frequently results in an ever-increasing interference problem as each additional station is placed in operation.

4. Single-frequency operation is most suitable where the operational need for single channel communication between mobile units (as contrasted with fixed-to-mobile) is more important than frequency economy.

5. A frequency separation between potential channel assignments of 25 kc in the 150 mc range, and 35 kc in the 450-mc range seems technically feasible; but adjacent channels with these minimum spacings cannot be assigned in the same area. These values may be reduced to about 20 and 30 kc, respectively, at the sacrifice of an appreciable reduction in coverage under impulse noise conditions. A further reduction in channel spacing would not appreciably increase the total number of usable channels, since the controlling factors are RF selectivity and extra band radiation, rather than IF selectivity or the total number of potential channel assignments.

6. The *average* spacing needed between channels operating in the same area varies from about 40 to 500 kc or more, depending on the method of operation and the criterion of usability.

7. The need is for a certain small number of channels in all areas, plus a peaked demand in centers of population. In the semi-coordinated cases, the maximum number of channels that can be allocated to the peak area is a small fraction of the total number of channels available. In the fully coordinated, broad-band case, there is much more flexibil-

ity and the peak area can be allocated a large fraction of the total available.

8. FM is preferable to AM for land mobile service because its instantaneous gain control feature minimizes the flutter caused by the motion of the mobile unit through standing wave patterns. This advantage increases in importance as the carrier frequency increases. In addition, FM with an adequate frequency swing provides an increased signal-to-noise advantage over most of the coverage area. The somewhat greater channel width required by FM is more than offset on an area coverage basis by the closer co-channel spacing.

Intermodulation Interference in Radio Systems

Frequency of Occurrence and Control by Channel Selection

By WALLACE C. BABCOCK

(Manuscript received August 25, 1952)

Intermodulation interference becomes a serious factor in frequency usage when a block of consecutive channels is provided for a given type of radio service in a confined area. Formulas are presented which show the number of potentially interfering 3rd and 5th order intermodulation products that can be formed in a band of n consecutive channels. The probability of encountering interference when a number of operating channels are picked at random from this band of n channels is developed and the number of interference free operating channels that can be obtained by careful selection in this same band is also derived.

When a block of consecutive radio channels is used in a confined area to provide a given type of service, interference becomes a serious problem. The situation is aggravated by the fact that whenever energy at two or more radio frequencies combines in a nonlinear circuit, as in transmitter output stages or in receiver input stages, products at other than the original frequencies are created. These are called intermodulation products, and they are capable of causing serious interference within the block of channels assigned to a given type of service as well as in other bands assigned to other types of service.

It is important in engineering a service to know something about the nature of these products in order to evaluate their interference potentialities and to study means of controlling or minimizing that interference. The numbers and locations of various types of intermodulation products are susceptible to mathematical computation. Whether or not all of these products would produce actual interference depends on the geographical locations of transmitters and receivers, and on their specific

electrical characteristics. This paper discusses the number of potential interferences, and in effect, envisages a situation where potential interferences are strong enough to be actual interferences.*

GENERAL

Intermodulation products are commonly referred to as 2nd, 3rd, 4th $\dots$ n th order products depending on the order of nonlinearity which gives rise to the products. Interference within a system is not generally experienced from the even order products because the frequency separation between the channels involved and the product formed by them is so great that the selectivity of the transmitter and receiver radio frequency circuits is sufficient to reduce it to a negligible amount. Some of the odd order products can be discounted also for the same reason. There are odd-order products, however, involving both sums and differences of operating frequencies in such fashion that the frequencies of the products formed are very close to those which generated them. These products are those referred to throughout the remainder of this paper since they are the most likely to cause interference. The most general form of 3rd order interference occurs when three frequencies, A , B and C , intermodulate in such fashion as to produce interference on a channel operating at frequency D . In this case

$$A + B - C = D$$

Another form of 3rd order interference occurs when the second harmonic of A intermodulates with B to produce interference on a channel operating at frequency C . In this case

$$2A - B = C$$

In like fashion the following forms of 5th order interference may occur.

$$A + B + C - D - E = F$$

$$2A + B - C - D = E$$

$$A + B + C - 2D = E$$

$$2A + B - 2C = D$$

$$3A - B - C = D$$

$$3A - 2B = C$$

* Bullington, K., Frequency Economy in Mobile Radio Bands. Page 42 of this issue.

NUMBER OF INTERMODULATION PRODUCTS AND FREQUENCY BAND AFFECTED

It is of interest to know how many intermodulation products can be produced by a block of n uniformly spaced channels and where they will fall with respect to the frequency band occupied by the n channels.

Third Order Products

When all possible combinations of n uniformly spaced operating channels are activated three at a time, $n^2(n - 1)/2$ products will be formed and they will lie between $A - (n - 1)$ and $A + 2(n - 1)$ where the operating band lies between channels A and $A + (n - 1)$. This means that the bandwidth of the intermodulation products is very nearly three times that of the operating channels. The products are symmetrically distributed with respect to the midpoint of the operating band. There will be

$$\frac{2n^3 + 3n^2 - 2n - 6}{24}$$

third order products that will fall in the $n - 1$ channels immediately below the operating band and a like number of products that will fall in the $n - 1$ channels immediately above the operating band. The remaining products,

$$\frac{4n^3 - 9n^2 + 2n + 6}{12}$$

in number, will fall in the n channels that constitute the operating band. Not all of these products, however, are capable of producing interference since some products of the $A + B - C$ type can fall on the very transmitting channels that combine to produce them. These products are harmless since they do not fall on receiving channels and are generally of much lower level than the carrier on the channels in which they do fall. If such products are not counted, there remain

$$\frac{n}{3} (n - 1)(n - 2)$$

products which fall within the operating band. The formulas presented here and in Table I are empirical and were derived for values of n up to 10. However, it is believed that they are reasonably accurate for much larger values of n .

Fifth Order Products

When all possible combinations of n operating channels are activated five at a time

$$\frac{n^2(n-1)(n^2-n+4)}{12}$$

fifth order products will be formed and they will lie between $A - 2(n-1)$ and $A + 3(n-1)$ where the operating band lies between channels A and $A + (n-1)$. This means that the bandwidth of the intermodulation products is very nearly five times that of the operating channels. The products are symmetrically distributed with respect to the midpoint of the operating band. It has been found empirically that if all products known to fall on transmitting channels are excluded, there can still be formed

$$\frac{1}{6}[6n^4 - 87n^3 + 575n^2 - 2214n + 3922]$$

potentially interfering products which fall within the operating band. This is strictly true only for values of n that exceed 8 and are not multiples of 3. There will remain

$$\frac{1}{12}[n^5 - 14n^4 + 179n^3 - 1154n^2 + 4428n - 7844]$$

products that will fall either outside the band or on transmitting channels within the band. Since relatively few of these products fall on transmitting channels within the band the above expression gives to a high degree of approximation the number of products that will fall outside the operating band.

Numbers of Potentially Interfering 3rd Order and 5th Order Products

The formulas given in the two preceding sections were developed by considering individually the various types of 3rd order and 5th order products. Table I shows the general formulas which apply to these individual types of 3rd and 5th order products from which the summation formulas given in the preceding sections were obtained. The number of potentially interfering 3rd and 5th order products is shown in Table II for specific transmission bands containing 10 and 20 consecutive channels.

PROBABILITY OF INTERMODULATION INTERFERENCE WHEN OPERATING CHANNELS ARE PICKED AT RANDOM

In an uncoordinated radio communication system it may be assumed that p operating channels are assigned on a random basis in a band con-

TABLE I — NUMBER OF POTENTIALLY INTERFERING INTERMODULATION PRODUCTS

 n = Number of consecutive channels in available band

Type of Product	Number of Products	
$2A - B$	$\frac{(n-1)^2}{2}$ if n is odd	$\frac{n(n-2)}{6}$ if n is even
$A + B - C$	$\frac{(n-1)(n-3)(2n-1)}{6}$ if n is odd	$\frac{n(n-2)(2n-5)}{6}$ if n is even
$3A - 2B$	$\frac{n}{3}(n-3)$ if n is a multiple of 3	$\frac{(n-1)(n-2)}{3}$ other n 's
$3A - B - C$	$3(n^2 - 11n + 32)$	
$2A + B - 2C$	$\frac{1}{2}(n^3 - 9n^2 + 34n - 56)$	
$2A + B - C - D$	$3(n-3)(n-5)^2$ for $n > 6$	
$A + B + C - 2D$	$4(n-3)(n-5)^2$	
$A + B + C - D - E$	$(n-5)(n-6)(n^2 - 11n + 37)$ for $n \neq 8$	

taining n consecutive channels. Let us suppose further that the average busy time of the channels is such that T represents the portion of time that an average channel is activated by a transmitter and R represents the portion of time that an average channel is connected to a receiver. The probability of interference, I , may be defined as the probability that one or more of the intermodulation products that are formed when pT channels are transmitting will fall on a specific channel in the operating band. The method used to determine I is based on the assumption that the distribution of intermodulation products is uniform over the operating band. It is further assumed that the magnitudes of the intermodulation products as encountered at the receiver input, are always strong enough to cause interference. Table III shows formulas for I that have been developed for each type of third order and fifth order product. Fig. 1 shows plots of p versus I when only 3rd order products are considered and Fig. 2 shows plots of p versus I when both 3rd and 5th order products are considered with n and T as independent variables.

1. Fig. 2a shows that a band in excess of 500 adjacent channels is required to limit the probability of 3rd and 5th order interference to 10 per cent ($I = 0.1$) when 10 operating channels are picked at random from that band if traffic is such as to fully use these 10 channels, 5 for

TABLE II — TYPE AND NUMBER OF INTERMODULATION PRODUCTS

Type of Product	Number of Products	
	10 Channels	20 Channels
$2A - B$	40	180
$A + B - C$	200	2,100
$3A - 2B$	24	114
$3A - B - C$	66	636
$2A + B - 2C$	192	2,512
$2A + B - C - D$	525	11,475
$A + B + C - 2D$	700	15,300
$A + B + C - D - E$	540	45,570

TABLE III — FORMULAS FOR I , PROBABILITY OF INTERFERENCE

n = Number of consecutive channels in available band.

p = Number of operating channels picked at random from those in band.

T = Portion of time average channel is activated by transmitter.

m = Number of intermodulation products of a specific type expected to fall within band of n channels when pT channels are activated.

Type of Intermodulation	Formula for m
$2A - B$	$m_1 = \frac{pT(pT - 1)(n - 2)}{2(n - 1)}$
$A + B - C$	$m_2 = \frac{pT(pT - 1)(pT - 2)(2n - 5)}{6(n - 1)}$
$3A - 2B$	$m_3 = \frac{pT(pT - 1)(n - 2)}{3n}$
$3A - B - C$	$m_4 = \frac{3pT(pT - 1)(pT - 2)(n^2 - 11n + 32)}{n(n - 1)(n - 2)}$
$2A + B - 2C$	$m_5 = \frac{pT(pT - 1)(pT - 2)(n^3 - 9n^2 + 34n - 56)}{2n(n - 1)(n - 2)}$
$2A + B - C - D$	$m_6 = \frac{3pT(pT - 1)(pT - 2)(pT - 3)(n - 5)^2}{n(n - 1)(n - 2)}$
$A + B + C - 2D$	$m_7 = \frac{4pT(pT - 1)(pT - 2)(pT - 3)(n - 5)^2}{n(n - 1)(n - 2)}$
$A + B + C - D - E$	$m_8 = \frac{pT(pT - 1)(pT - 2)(pT - 3)(pT - 4)(n - 5)(n - 6)(n^2 - 11n + 37)}{n(n - 1)(n - 2)(n - 3)(n - 4)}$

$I = 1 - \left(1 - \frac{1}{n}\right)^{m_1 + m_2}$ when only third order products are considered.

$I = 1 - \left(1 - \frac{1}{n}\right)^{m_1 + m_2 + \dots + m_8}$ when both third and fifth order products are considered.

transmitting and 5 for receiving. ($T = 0.50 = R$) In this case it is the $2A + B - 2C$ type of product that requires the use of such a wide band.

2. Fig. 2a shows that there is practical certainty of interference in a band of 500 adjacent channels when 30 operating channels, picked at random from that band, are fully used. In this case it is the $A + B + C - D - E$ type of product that requires the use of such a wide band.

3. If the same total traffic as was assumed in (1) is handled by a greater number of operating channels, the number of available consecutive channels required for the same probability of interference remains the same. Thus, Fig. 2b shows that a band in excess of 500 consecutive channels is still required to limit the chance of interference to ten per cent when the traffic is distributed among 20 randomly selected operating channels; ($T = 0.25 = R$) similarly Fig. 2c shows that the required number of available consecutive channels remains the same when the traffic is distributed among 40 operating channels. ($T = 0.125 = R$).

CHANNEL SELECTION FOR THE ELIMINATION OF INTERMODULATION INTERFERENCE

Discounting the effect of selectivity in the radio equipment, it was shown in the preceding section that only a very limited number of channels can operate together without some degree of mutual interference when these channels are picked at random from a very considerable number of available channels. This is of course extravagant of frequency space. In this section, it is proposed to determine whether frequency space can be conserved by carefully selecting the operating channels in such fashion that the various types of intermodulation products that are formed will all fall on other than operating channels. This is readily accomplished in the case of 3rd order products by selecting the operating channels in such fashion that the frequency difference between any pair of these channels is unlike that between any other pair of channels. Many other rules inherently more complicated and more cumbersome to apply than the one stated above must be obeyed if 5th order as well as 3rd order products are to be controlled in this way. Table IV presents p operating channels selected from a band of n adjacent channels (numbered sequentially in order of ascending frequency) in such fashion as to avoid 3rd order interference within the system. Considerable effort has been spent in selecting these channels to insure that the number n associated with each value of p is the lowest

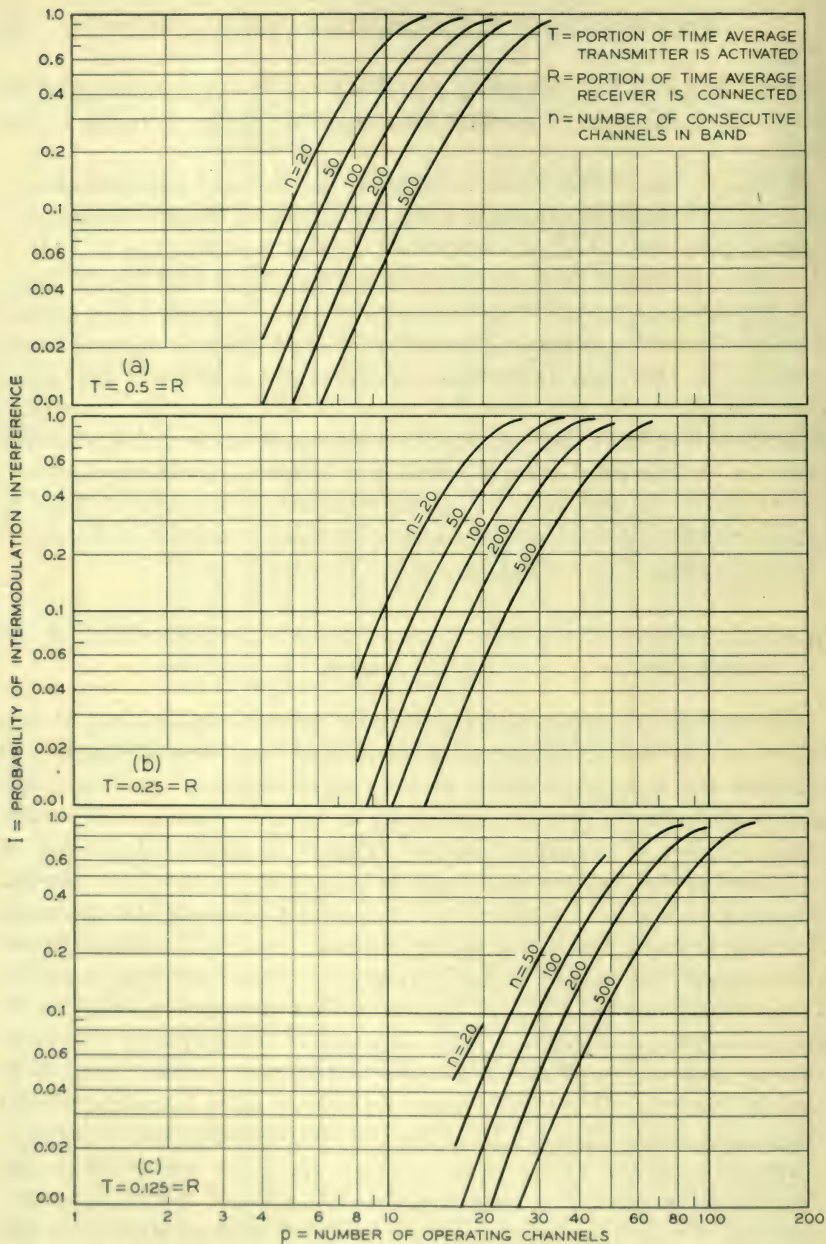


Fig. 1—Probability of intermodulation interference I versus number of operating channels p when only 3rd order products are considered. (a) T = Portion of time average transmitter is activated = 0.50; R = Portion of time average receiver is connected = 0.50. (b) Same as above except $T = 0.25 = R$. (c) Same as above except $T = 0.125 = R$.

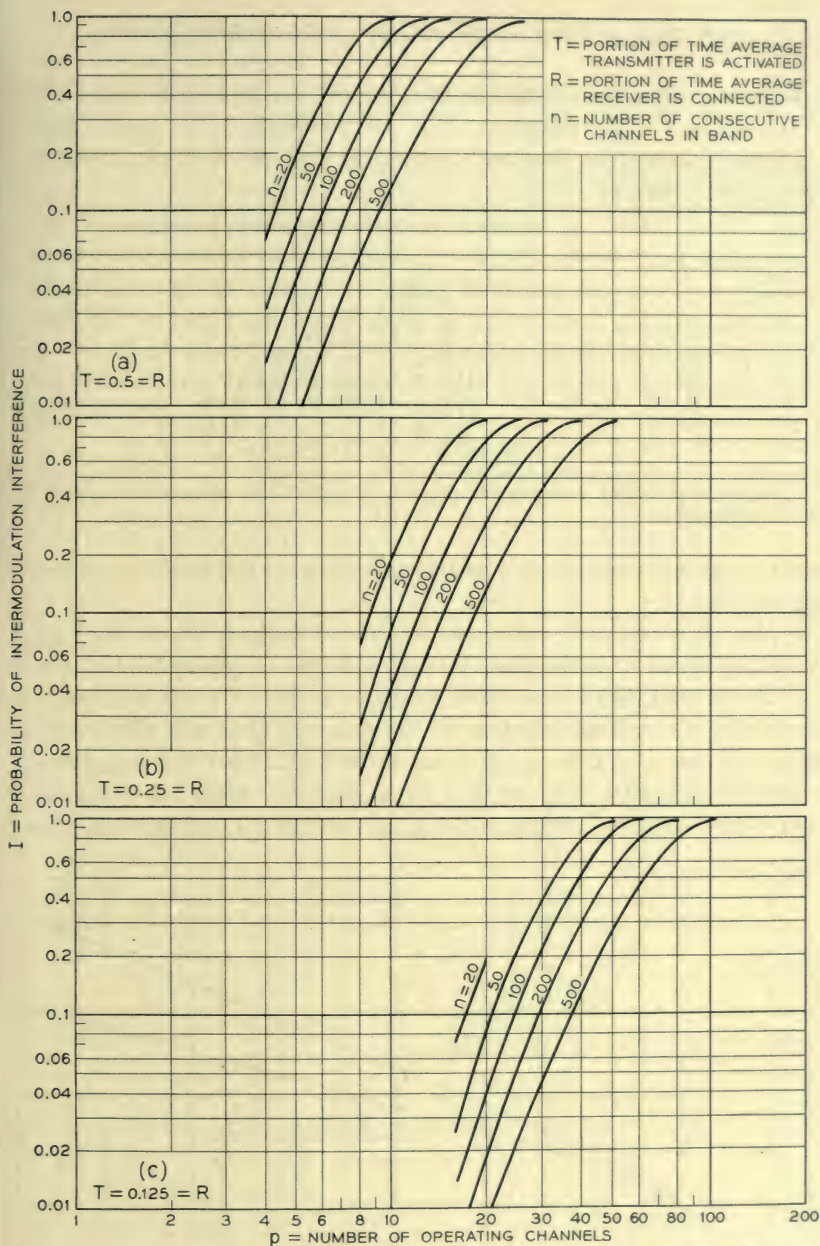


Fig. 2—Probability of intermodulation interference I versus number of operating channels p when both 3rd and 5th order products are considered. (a) T = Portion of time average transmitter is activated = 0.50; R = Portion of time average receiver is connected = 0.50. (b) Same as above except $T = 0.25 = R$. (c) Same as above except $T = 0.125 = R$.

TABLE IV — SPECIFIC OPERATING CHANNELS HAVING NO THIRD ORDER INTERMODULATION INTERFERENCE

p is the number of interference free operating channels which can be obtained from n consecutive channels.

p	n	Operating Channels Having No 3rd Order Interference
3	4	1, 2, 4
4	7	1, 2, 5, 7
5	12	1, 2, 5, 10, 12
6	18	1, 2, 5, 11, 13, 18
7	26	1, 2, 5, 11, 19, 24, 26
8	35	1, 2, 5, 10, 16, 23, 33, 35
9	46	1, 2, 5, 14, 25, 31, 39, 41, 46
10	62	1, 2, 8, 12, 27, 40, 48, 57, 60, 62
8*	137	1, 2, 8, 12, 27, 50, 78, 137

* Neither 3rd nor 5th order interference exists with this selection of eight operating channels.

possible number from which p channels having no 3rd order interference can be selected.

A plot of p versus n based on the above table is shown in Fig. 3. The curve which includes both 3rd and 5th order intermodulation products shows that 300 consecutive channels must be made available to provide for the careful selection of 10 operating channels which are interference-free at all times, regardless of the traffic load. For comparison, it was shown earlier (Fig. 2) that more than 500 consecutive channels must be available to permit picking at random 10 operating channels

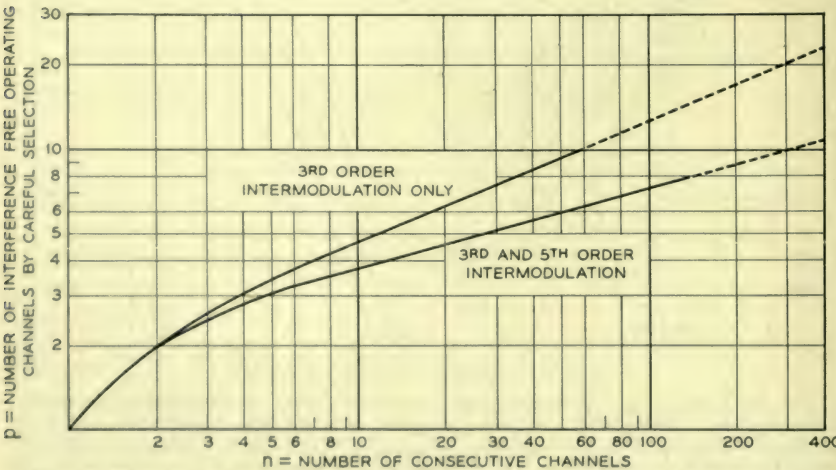


Fig. 3—Number of consecutive channels n required to provide a number of interference free operating channels p .

which are subject to interference 10 per cent of the time when fully loaded with traffic.

Careful channel selection is therefore a step toward better frequency usage. It is, however, only a small step, and a more effective solution must be found if efficient frequency usage is to be achieved in areas where sizeable numbers of operating channels are required. Such a solution for the mobile radio service is a "coordinated system," proposed in a companion paper previously referred to, wherein additional measures are described for reducing the probability of interference by proper geographical system layout and the use of certain practicable operational features.

Magnetic Resonance

PART I—NUCLEAR MAGNETIC RESONANCE

By KARL K. DARROW

(Manuscript received September 3, 1952)

Magnetic resonance is the name of a phenomenon discovered less than sixteen years ago, which from the start has had a high theoretical importance and is now attaining a notable practical value. Nuclear magnetic resonance occurs when a substance containing magnetic nuclei is exposed to crossed magnetic fields, one being steady and the other oscillating, and the strength of the former field and the frequency of the latter are matched in a particular way. When these are properly matched, the nuclei are turned over in the steady field, and energy is absorbed from the oscillating field. Another way of describing the effect is to say that resonance occurs when the applied frequency is equal to the frequency of precession of the nuclei in the steady field. This phenomenon illustrates very clearly some of the fundamental laws of Nature. For the purposes of nuclear physics it is used to determine the magnetic moments of nuclei and their relaxation-times in the substance that contains them. It is also used for chemical analysis, for measurement of magnetic fields, for analysis of crystal structure and for locating changes of phase of the substance containing the nuclei. Magnetic resonance of electrons is similar, but for a fundamental reason is confined almost exclusively to free atoms of certain kinds, to ferromagnetic substances and to certain strongly paramagnetic salts. For these last it serves to throw light on the fields prevailing within the crystals.

“Magnetic” is an ancient word in physics and so is “resonance,” but “magnetic resonance” is something new. It is the name of a phenomenon which is sharp and clearcut and easy to evoke, which springs directly from the ultimate magnetic particles of matter, which illustrates the fundamental laws of these, and which has found and still is finding uses of importance. There are two types of it, the nuclear and the electronic. Nuclear magnetic resonance is the theme of the first part of this article: it will recur from time to time in the second part (to appear in a later issue of this JOURNAL) but the main topic of that second part will be

electronic resonance. It is fitting that they should be treated in this order, for the nuclear type of resonance is less distorted by complexities than is the other. Perhaps it is not premature to say that while nuclear magnetic resonance always goes by that name, the electronic type is usually called "paramagnetic resonance" or "ferromagnetic resonance."

Nuclear magnetic resonance was realized in 1937, in molecular-beam experiments. The war distracted physicists, and the next great step was not made until after — but very soon after — the armistices. In the winter months of 1945–46 the phenomenon was produced in liquids and in solids. The news burst upon the world from the pages of *The Physical Review* in the early weeks of 1946, causing among physicists an immediate and an immense sensation. Of some discoveries one wonders how they came to be made at all, of others one wonders afterward why they were not made earlier. Nuclear magnetic resonance is of the latter class. But this is a discovery that could not have been made *much* sooner than it was, for it required the apparatus and techniques of short-wave radio and microwaves, and these are recent.

The work of 1945 was done by two independent groups three thousand miles apart, using somewhat different experimental methods and expounding the theory in somewhat different ways. The differences are really superficial, and in the course of time will probably be minimized; but the two streams of later work that rose from those two sources are still distinguishable. The methods are called the nuclear resonance absorption method and the nuclear induction method: I treat them in this order. A sketchy account of the molecular-beam method will follow upon these, and then several of the applications — which of these are major and which are minor must be left for history to decide.

On the first few pages, and on many thereafter, the talk will be of protons. Protons are the commonest material particles in Nature, electrons excepted (neutrons are also an exception but not an important one here, as they are seldom found free). Protons also have the happy attribute called "spin $\frac{1}{2}$ " soon to be explained, which simplifies the exposition greatly. This is one of the rare fields of physics in which the simplest case, the commonest case, and the most useful case, are all three of them one and the same.

PROTON RESONANCE ABSORPTION

To begin with, there must be a sample containing hydrogen, protons being the nuclei of ordinary (as distinguished from heavy) hydrogen atoms. It may be pure hydrogen in gaseous, liquid or solid form, or any

one of countless compounds of hydrogen. I first take water for the sample, and enumerate the other particles in water. There are the nuclei of the common isotope of oxygen, oxygen 16: they are non-magnetic and produce no resonance. There are the nuclei of rarer isotopes of oxygen and hydrogen: they will be mentioned later. There are the electrons: they are reserved for Part II of this article. We are now left with the protons.

The sample is placed between the poles of a magnet, Fig. 1, so that it is in a magnetic field which should be homogeneous and is usually strong. The strength of the field is denoted by H , and its direction is always that of the z -axis (and usually vertical). I should like to call it "the steady field," but usually it is modulated during the experiments, so I shall call it "the big field". Actually it can be very small, but nearly always it is between 8,000 and 15,000 gauss, and 10,000 gauss is a good figure to keep in mind.

The big field must not be the sole magnetic field applied to the sample. There must also be an alternating or oscillating field — stationary electromagnetic waves, formed in a solenoid (or sometimes in a resonant cavity). Such waves comprise, as Maxwell taught us long ago, an alternating electric field and an alternating magnetic field. In most of the uses of electromagnetic waves it is the electric field that counts, and the magnetic field is remembered only as something demanded by Maxwell's equations to keep the electric field going. In this application the electric field takes a back seat, and it is the magnetic field that counts. This oscillating magnetic field must be at right angles to the big field; we lay the x -direction along it. Its amplitude, to be denoted

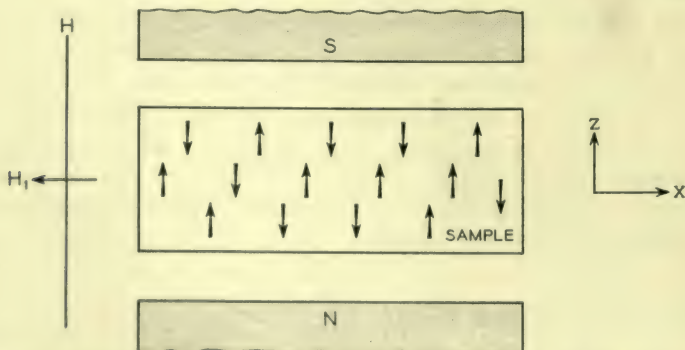


Fig. 1 — Scheme of the apparatus for observing nuclear magnetic resonance. The detecting circuits are omitted. The nuclei indicated by the arrows are of "spin $\frac{1}{2}$," protons for example.

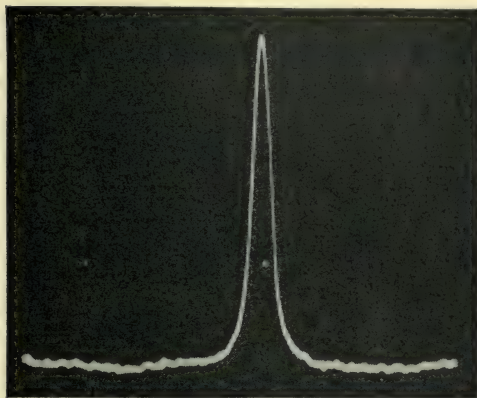


Fig. 2 — Peak of nuclear resonance absorption. This is the first peak to be published other than those obtained with molecular beams. It pertains to protons in water. (Courtesy of E. M. Purcell).

by H_1 , is of the order of a small fraction of one gauss up to several gauss. Its frequency must be of the order of tens of megacycles. To be more specific, the effect that is sought with protons is located at 42.6 mc when H is 10,000 gauss.

Finally there must be circuits and detectors for measuring the absorption of the electromagnetic wave-energy in the sample. These are well known to those proficient in the art: we pass them over.

Now of the two quantities H and ν either is to be varied while the other is to be kept constant, and the absorption is to be measured. Usually H is varied while ν is kept constant, and the data consist of a plot of absorption against H for a set value of ν .

When such a curve is plotted it proves to be, in the main, a smoothly-sloping curve, of no interest in the present connection. What *is* of interest is that it is interrupted by a magnificent peak of extraordinary sharpness, deserving to be called a needle. Probably there is nothing that can please an experimenter more than a curve with a fine sharp peak: here he has it. Fig. 2 exhibits the first such peak on record. But neither Fig. 2 nor any other picture can convey an adequate idea of the sharpness of the peak, for the distance from this imposing feature to the axis of the ordinates at field-strength zero may be, and often is, tens of thousands of times as great as the breadth of the peak. (With the induction-method, peaks have been distinguished from each other that are separate by only a millionth of the value of H at which they are found). This needle has the narrowness that is characteristic of fine lines in optical spectra; and this is as it should be, for a spectrum-line

is just what it is, even though it lies in the radiofrequency range and for technical reasons appears on a scope instead of on a photographic film.

The peak is the phenomenon of magnetic resonance. We shall now interpret it in terms of theory somewhat oversimplified, for it is not the office of these opening paragraphs to introduce all of the complexities of quantum mechanics.

In Fig. 1, inside the rectangle which represents the sample, appear a number of arrows. These are symbols of the protons. In other circumstances we might imagine the protons as solid balls, in still others we might imagine them as centres of force; but for the present purpose we are regarding them as tiny bar-magnets, and the arrows symbolize the directions in which they are pointing. It is necessary to label these directions with perfect clearness. The figure has been drawn with the south pole-piece of the big magnet (the one responsible for the big field H) above. The point of each arrow represents the *north* pole of the protonic magnet.

Thus the arrows pointing upward represent protons in the orientation into which the big field would like to turn the protonic magnets, and would indeed succeed in turning them if they were literal compass-needles in literal compasses. The arrows pointing downward represent protons in the opposite orientation. I will call these, for shortness, the "up" orientation and the "down" orientation. Evidently if the physicist could reach into the sample with fingers or with forceps and turn a proton from the up orientation into the down one, he would be doing work upon the proton at the expense of energy from his muscles. Well, he cannot reach into the sample with fingers or with forceps and grasp and turn a proton. But he can reach into the sample with the oscillating field and turn the protons, and this is the experiment we are considering. Magnetic resonance is the turning of protons from the up orientation into the down one, from the orientation or "level" of lesser energy into the orientation or level of greater energy.

But why does the effect occur at one frequency only? And what determines that frequency? To cope with this problem we shall have to introduce symbols, equations, and quantitative reasoning.

The first step is to evaluate the work required to turn the proton, or, in other words, the energy-difference between the two orientations or levels. It shall be denoted by W , and the magnetic moment of the proton by μ_p . We proceed by strictly classical reasoning. The torque exerted on the proton by the magnetic field H is $-\mu_p H \sin \theta$. Here θ stands for the angle between the direction of the steady field and the direction in which the magnetic moment of the proton is pointing. We have

admitted the existence of only two values of θ , viz. the values 0° and 180° ; more will be said about this later; but for the duration of this particular argument we shall have to admit all values of θ from 0° to 180° . The value of W which we are seeking is the integral of $\mu_p H \sin \theta$ from 0° to 180° , from the up orientation to the down one. It is easily obtained:

$$W = - \int_{0^\circ}^{180^\circ} \mu_p H \sin \theta d\theta = 2\mu_p H \quad (1)$$

Having arrived at equation (1) by strictly classical reasoning, we must now approach equation (2) by a starkly quantal argument. Immense amounts of evidence have shown that when energy is absorbed from electromagnetic waves of frequencies ν in the optical range of the spectrum and in the X-ray range, not to speak of other ranges, it is invariably absorbed in parcels or quanta equal to $h\nu$, h standing as always for Planck's constant. If this doctrine is sometimes difficult to assimilate when applied to the optical spectrum, how much more difficult it is to accept when applied to waves of radio frequencies! Yet here also it is to be accepted, so we put:

$$W = h\nu \quad (2)$$

Now we transfer the value of W from equation (1) to equation (2), and arrive at the destination:

$$H = \frac{1}{2}h\nu/\mu_p \quad (3)$$

In this equation h is known with very great accuracy, and μ_p had also been measured when the first experiments upon magnetic resonance were made, though not with nearly the accuracy that physicists now claim for it. It remains only for the experimenter to insert for ν the value of the frequency in his experiment and for H the value of the fieldstrength at which the peak appears. The test is whether the two sides of the equation agree. Needless to say, the test has been brilliantly passed.

Quantum-theory has entered into this argument in more ways than the one which led to equation (2). I return now to the fact that we have arrived at equation (3) by postulating two, and only two, "permitted" orientations of the protonic magnets in the steady field. This is illustrated by the presence, in Fig. 1, of arrows pointing up and arrows pointing down but no arrows pointing slantwise. We might have assumed that there are protons, and therefore arrows, pointing in every direction. We might have assumed that there is a proton pointing, say, at angle

$76^{\circ}13'$ to the vertical, and that it can absorb a quantum $h\nu$ of just the right energy to turn it to the angle $118^{\circ}36'$. This would have led to the inference that instead of absorption confined to the fieldstrength $h\nu/2\mu_p$ corresponding to the actual peak, there would be absorption at every fieldstrength from $h\nu/2\mu_p$ on upwards toward infinity. The experiment frustrates this inference, and so declares for the two and the only two permitted orientations. One did not have to wait for this experiment to learn this fact: it has been known for thirty years, both as a consequence of quantum mechanics and as a fact of experience. However this is a very pretty proof of it.

We now must generalize equation (3) so as to make it take care of all nuclei and not the proton only; and in the course of this process we shall meet the actor behind the scenes who determines the permitted orientations. His names are *spin* and *angular momentum*.

THE GENERAL EQUATION FOR NUCLEAR MAGNETIC RESONANCE

We are now en route to the general equation of which (3) is the special case appropriate to the proton. Our first step takes us to the deuteron or nucleus of heavy hydrogen. Its magnetic moment differs from that of the proton, so we must write μ_d instead of μ_p . More significant is the fact that the deuteron has three permitted orientations in the big field instead of two. The orientations of proton and deuteron are shown in the first and third columns of Fig. 3; beside them are horizontal lines depicting their energy-values, energy being measured vertically upward from an arbitrary zero.

One guesses from the aspect of Fig. 3 that the deuteron will show three peaks of magnetic resonance; for it seems possible for the deuteron to be turned from orientation *a* to orientation *b*, from *b* to *z* and from *a* all the way to *z*. But of these three conceivable "transitions" the third

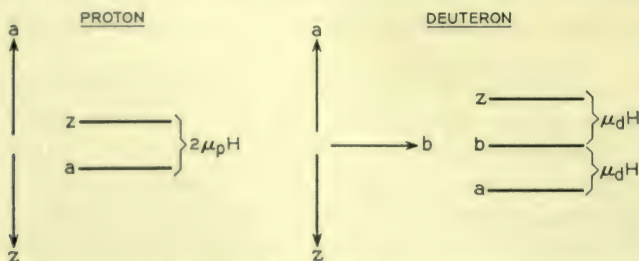


Fig. 3 — Orientations and energy-levels of protons and deuterons in a magnetic field, according to the "old" quantum-theory.

does not occur at all: the theorists know the reason why, and call it a "forbidden" transition. As for the other two, the mere symmetry of the picture shows that they involve equal absorptions of energy and therefore contribute coincident peaks. There is therefore only one distinguishable peak, and we have to find the value of H at which it appears. This is easily done. Going back to equation (1), we integrate the integrand which there appears from 0° to 90° (or from 90° to 180°); and so we come to the analogue of equation (3) which applies to the deuteron:

$$H = h\nu/\mu_d \quad (4)$$

In the course of this argument we have met with an example of two general rules: *no matter how many permitted orientations there are, transitions occur only between consecutive ones, and these permitted transitions always agree in energy-absorption, so that there is never more than one peak.* Yet equation (4) differs from equation (3), because in the right-hand member $h\nu/\mu$ — and now I am using μ as the general symbol for magnetic moment — is multiplied by $\frac{1}{2}$ for the proton and by one for the deuteron. Now, $\frac{1}{2}$ is the value of the spin of the proton and one is the value of the spin of the deuteron. We generalize from these two instances: we use I as the general symbol for the spin; and we arrive at the following:

$$H = (I/\mu)h\nu \quad (5)$$

The generalization is sound; and *equation (5) is the fundamental equation of magnetic resonance.*

I now have to interpret the word "spin." Spin is a particular measure of the angular momentum of the nucleus. That a magnetic nucleus has angular momentum is surely not surprising. We are trained to ascribe magnetism to the motion of charged bodies: an electric current flowing in a loop has the same magnetic field as a bar-magnet. When a nucleus is observed to have a magnetic moment and an angular momentum, it is natural to correlate one property with the other: one does not quite know how far the analogy may safely be pressed, but at least it is helpful.

But what sort of a measure of the nuclear angular momentum is the quantity I ? The answer to this question is confused by the fact that in our times there have been two forms of quantum theory: the "new" quantum mechanics which is undoubtedly more competent in general, and the "old" quantum theory of the nineteen-twenties which is certainly more simple in the present case. Desire to be clear has led me to employ

the older theory up to now, but conscience obliges me to introduce the new one.

In the old theory, I is the nuclear angular momentum in terms of the unit $\hbar/2\pi$. Two of the permitted orientations, which I will call the "extreme" ones, are straight along and straight against the field-direction. For these, the projections of the angular momentum upon the field-direction are $+I\hbar/2\pi$ and $-I\hbar/2\pi$. For the proton $I = \frac{1}{2}$, and the two extreme orientations are the only ones.

In the new theory, the nuclear angular momentum in terms of the unit $\hbar/2\pi$ is $\sqrt{I(I+1)}$. For the two extreme orientations, the projections of the angular momentum on the field-direction are $+I\hbar/2\pi$ and $-I\hbar/2\pi$, just as they were in the old theory. But now these orientations are no longer straight along and straight against the field-direction. They must be inclined, one to the up direction and the other to the down direction, at the angle of which the cosine is $I/\sqrt{I(I+1)}$.

Thus I has partly changed its meaning: it is still the maximum permissible projection, upon the field-direction, of the nuclear angular momentum in terms of the unit $\hbar/2\pi$, and this is what it was before; but it is no longer the magnitude of the nuclear angular momentum. So also has μ changed a part of its meaning. It is the maximum permissible projection, upon the field-direction, of the magnetic moment of the nucleus, and this it was before; but it is not the magnitude of the nuclear magnetic moment. The language of this subject has not been well adjusted to this change. Fortunately I is called the "spin," which does not necessarily convey the impression that it is quite the same thing as angular momentum; but μ is still called the "magnetic moment," and in the new quantum mechanics this is a mistake.

Fig. 4 is Fig. 3 redrawn in the spirit of quantum mechanics. The arrows now represent angular momentum and magnetic moment jointly,

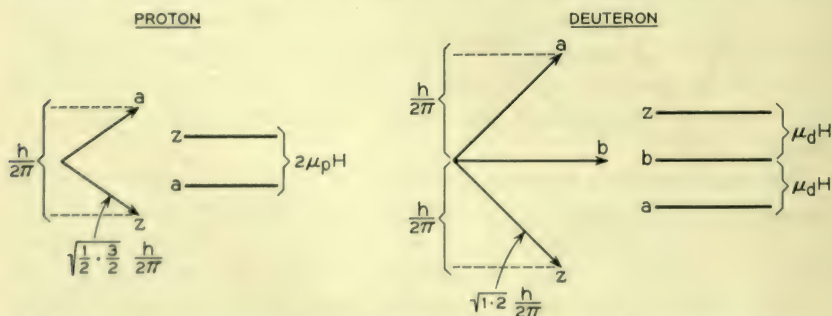


Fig. 4 — Orientations and energy-levels of protons and deuterons in a magnetic field, according to quantum mechanics.

but the numbers affixed to them are the values of angular momentum. The energy-levels in the second and fourth columns are often known as "Zeeman levels." I take this occasion to complete the statement about the allowed orientations, which in recent paragraphs has been made for the extreme orientations only. The projections of the nuclear angular momentum upon the field-direction are

$$+Ih/2\pi, +(I-1)h/2\pi, \dots, -(I-1)h/2\pi, -Ih/2\pi.$$

From this principle combined with the fact that transitions occur only between consecutive levels, follows rigorously equation (5), which I derived in a looser way.

Spins are ascertained in various ways, usually from their influence on the electrons surrounding the nuclei, which manifests itself in details of optical spectra and in cleverly-designed molecular-beam experiments. They are always integer multiples of $\frac{1}{2}$. Important instances of nuclei of spin $\frac{1}{2}$ are the proton and the nucleus F^{19} . The neutron and the electron also belong in this category, as we shall see later on. The deuteron has already provided us with an important instance of a nucleus of spin one. Spins as high as $\frac{9}{2}$ are certainly known, and this is probably not the limit. Nuclei of spin zero are common: I have already mentioned one of them, oxygen 16. Such nuclei do not produce magnetic resonance; we shall have nothing to do with them.

A brief table shall conclude this section. To what has already been stated it adds the number of permitted orientations corresponding to each value of spin.

Spin.....	$\frac{1}{2}$	1	$\frac{3}{2}$	I
Number of orientations.....	2	3	4	$2I + 1$
H for peak.....	$(\frac{1}{2})h\nu/\mu$	$h\nu/\mu$	$(\frac{3}{2})h\nu/\mu$	$(I/\mu)h\nu$

THE LARMOR PRECESSION AND NUCLEAR INDUCTION

Now we go back to first principles, make a fresh start, and arrive by a different route at the equation for magnetic resonance. On this route we meet with a vivid justification of the use of the name "resonance."

Resonance implies a tuning or a matching between an applied frequency and a frequency either actually or potentially present in the substance in question. A piano-wire, a membrane, the air in an organ-pipe, an electrical circuit comprising capacity and inductance, all resonate to the frequency which is that of their own natural vibrations. No mention has yet been made of a frequency peculiar to the nucleus which is matched by the applied frequency when magnetic resonance

occurs. There is indeed such a natural frequency, not however a frequency of vibration; it is a frequency of *precession*. Precession is a concept well known to astronomers and to such physicists as have to do with gyroscopes, perhaps not so well known as it should be to others.

In Fig. 5, the vertical is again the direction of the big magnetic field. The arrow represents the angular momentum of the nucleus, which I now denote by p . The magnetic field H exerts a torque on the nucleus. I have already given an expression for this torque, but I gave it in the language of the "old" quantum-theory. To employ this expression with as little apparent change as possible, I introduce the symbol μ_0 for the magnetic moment of the nucleus, and reserve μ for

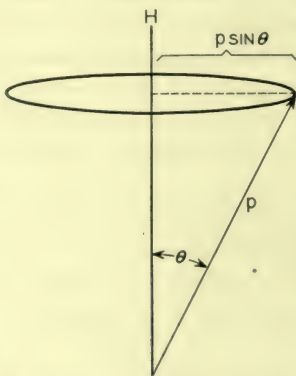


Fig. 5 — Illustrating the Larmor precession.

the maximum permissible projection of μ_0 on the field-direction. The torque now appears on the right-hand side of the following, purely classical, equation:

$$dp/dt = \mu_0 H \sin \theta \quad (6)$$

This is a vectorial equation, but I will endeavor to express its vectorial content by words instead of symbols. Fix the attention on the tip of the arrow. The torque makes it describe a circle of radius $p \sin \theta$ in the horizontal plane, with a frequency which I denote by ν . Its peripheral speed in this circle is ν multiplied by the circumference of the circle, therefore $\nu \cdot 2\pi p \sin \theta$. This speed is dp/dt . Putting its value into (6), we observe with pleasure that θ vanishes from the scene: the result is going to be the same for all orientations of the magnet: this is it:

$$H = 2\pi p \nu / \mu_0 \quad (7)$$

Making the substitutions that have already been describe, we get:

$$H = (I/\mu)h\nu \quad (8)$$

and this is none other than formula (5), the fundamental equation of magnetic resonance. The precession-frequency is the resonance-frequency.

This precession is often called the "Larmor precession," and the frequency given by (5) or (8) is called the "Larmor frequency." The name is a posthumous honor; Larmor died before magnetic resonance was discovered; his theory was applied to the Zeeman effect, the effect of magnetic fields upon optical spectra.

It is not hard to believe that when the applied frequency coincides with the Larmor frequency, something drastic must happen to the precession. The theory has been worked out on a classical basis. I will not pursue it into its details; but at least the first step should be taken.

I have said that the alternating field is perpendicular to the big field. We take the x -direction as its direction. The magnetic field, or magnetic vector as I will henceforth call it, has then $H_1 \cos 2\pi\omega t$ for its x -component (I use ω for the frequency so as to distinguish it from the Larmor frequency) and zero for its y -component. Now imagine a vector, of constant magnitude $(1/2)H_1$, lying in the xy -plane, pointing away from the z -axis and revolving around this axis *clockwise* with frequency ω . Its x -component will be $(1/2)H_1 \cos 2\pi\omega t$, its y -component will be $-(1/2)H_1 \sin 2\pi\omega t$. Imagine another such vector revolving *counterclockwise*. Its x -component will be $(1/2)H_1 \cos 2\pi\omega t$, its y -component will be $(1/2)H_1 \sin 2\pi\omega t$. (It is evident that we have chosen their phases so as to bring about this result). The sum of these two vectors has $H_1 \cos 2\pi\omega t$ for its x -component and zero for its y -component. But this is the vector that we started out with. In the language of optics, we have resolved a plane-polarized wave into two circularly-polarized ones.

The foregoing is pure mathematics. Now comes the physics. Of these two revolving vectors, one is whirling in exact unison with the precessing magnet when ω is exactly equal to the Larmor frequency, the other is rushing round and round in the opposite direction. Our intuition tells us that the former may be expected to produce a great effect on the precession, the latter a small one. The latter is not always negligible, but may be neglected here. Thus in this artful way we have substituted a circularly-polarized field for the actual plane-polarized one.

The theory further leads to the prediction that when resonance exists, the precession will be exaggerated in such a way as to produce

an alternating magnetic flux across the xz plane. Now I describe an actual experiment, the first of its type.

The sample is water (or something else) in a spherical container. Around the container are wrapped two coils at right angles to one another. The coil of which the axis is parallel to the x -axis produces the alternating field. The coil of which the axis is parallel to the y -axis is connected with a rectifier and a detector. At resonance there is an alternating magnetic flux through the latter coil, and by the operation of the rectifier this is converted into a signal on the scope. The signal locates the resonance-frequency as accurately as does the peak in the absorption-method. This is the phenomenon called "nuclear induction."

I terminate this section by mentioning a paradox resulting from precession. Everyone has seen a compass-needle turning to point to the north: it is natural to infer that when a magnetic field is applied to a piece of matter, the elementary magnetic particles of which the nuclei (and also the electrons) are examples will automatically turn to point along the field. Yet the analogy fails and the inference is false: the nuclei do not turn to point along the field, but each of them maintains a constant angle with the field while it precesses. It seems to follow that matter cannot be magnetized by a magnetic field, but again the inference is false. Animistically speaking, the field makes the nuclei want to turn into its direction, but they cannot fulfill their desire without assistance from something other than the field. This something-other is not absent, and in the section on "relaxation" we shall meet with it.

THE MOLECULAR-BEAM EXPERIMENT

There are three methods for detecting and locating nuclear magnetic resonance, and we have now considered two of them. In one of these, the resonating nucleus makes itself manifest by absorbing energy; in the other, that of nuclear induction, by radiating energy; in the one which is to come, by simply failing to turn up at the scene of the measurement. This singular attribute is that of the molecular-beam experiment, which (I repeat) was done before the others and so receives the credit of revealing nuclear magnetic resonance. Molecular-beam experiments are so remarkable that it is hard to speak of them without yielding to temptation to say more than is essential to the purpose, but here the temptation must be withstood.

Conceive a narrow stream of hydrogen-containing molecules coming along the (horizontal) axis of y , and cutting across a big magnetic field parallel to the (vertical) axis of z . This big field differs from that of

Fig. 1 in one important way: it is *non-uniform*, increasing in strength from (say) the bottom to the top. In one respect the protons behave just as they do in a sample in a uniform field: roughly half of them are pointing up and the other half pointing down. But in the non-uniform field the "up" protons experience a net force pushing them upward and the "down" protons a net force pushing them down. (Visualizing each of the protons as a tiny bar-magnet, one sees that the field strength is bigger where the upper pole of the magnet is than where the lower pole is). The beam is parted into two diverging pencils, the one containing the "up" protons only and the other the "down" protons only; I call the first the "up" pencil and disregard the second.

The "up" pencil now passes through a region just like that implied in Fig. 1: a magnetic field which is big and vertical and *uniform*— H will stand for its strength—and an oscillating field with the magnetic vector parallel to the x -direction. If in this second region some of the protons are turned by the oscillating field into the "down" orientations, that will make no difference to their course across the remainder of the second region where H is uniform. But beyond the second region lies a third where again there is a big field that is non-uniform. In this third region the "up" protons go one way and the "down" protons go another. The detector lies athwart the first way; the "down" protons will miss it.

The detector-reading is plotted against H for a set value of ν . One might think that two curves would be plotted, one with the alternating field off and the other with it on, and that the latter would be systematically lower than the former. But the latter will be lower than the

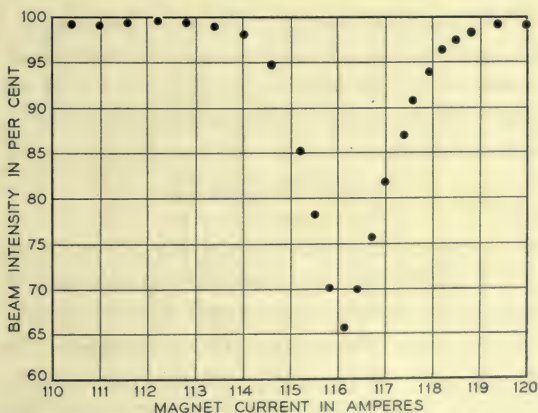


Fig. 6 — Negative peak or valley of nuclear resonance absorption obtained by the molecular-beam method. It pertains to lithium nuclei in lithium chloride molecules. This was the first experimental evidence of nuclear magnetic resonance. (I. I. Rabi, J. R. Zacharias, S. Millman and P. Kusch).

former only in the immediate vicinity of the value of H which conforms to equation (3); for the protons are turned over only when the Larmor frequency agrees very nearly with the applied frequency. Accordingly one keeps the alternating field on all the time and plots a single curve; and this is marked by a fine sharp peak, but this time a peak that points *downward*, Fig. 6, for it testifies to the absence of the overturned protons that have missed the detector.

The first experiment of this kind was done on molecules of lithium chloride. The reader may have been puzzled that I spoke of a beam of *molecules* and then of the deflection of *protons*: the protons, or whatever other magnetic nuclei are being studied, carry the molecules with them. In the experiments on LiCl, the peaks of lithium and of chlorine were found in different parts of the curve. Later the proton-resonance was discovered by using molecules of KOH and NaOH, and confirmed with molecules of H_2 and HD (the latter being a hydrogen molecule of which one nucleus is a proton and the other a deuteron). It is from this molecule of HD that the proton-resonance, and for that matter the deuteron-resonance also, stand out most clearly and sharply. In H_2 and in D_2 the resonances are perturbed and multiplied, but for reasons which are well understood so that the theory is strengthened instead of being weakened; but to describe these pretty things would be confusing unless they were explained, and to explain them would take us far afield.

SOME APPLICATIONS OF NUCLEAR MAGNETIC RESONANCE

The first of the uses of nuclear magnetic resonance is of interest mainly to the nuclear theorist. He wants to know (I/μ) for as many nuclei as possible; and this knowledge may be found by locating the resonance-peaks, and applying to their values of H and ν the equation (3) or (5) which I repeat:

$$H = (I/\mu)h\nu \quad (9)$$

Anyone who is going to burrow into the literature of this subject must be apprised beforehand, or else find out the hard way, that this simple statement is variously expressed. Here is a sad case of the ruination of a beautiful terminology by carelessness. The terms which have been ruined are "gyromagnetic ratio" and "magneto-mechanical ratio." The former ought to mean, as originally it did mean, the ratio of angular momentum to magnetic moment. The latter ought to mean the ratio of magnetic moment to angular momentum. Both have by

now been used in both these senses, and there are variants within each sense, depending on the unit that is preferred by the user. The appearance of either "gyromagnetic ratio" or "magneto-mechanical ratio" in a paper is a red light warning the reader to make sure just what the author means by it. In this paper both of these terms are discarded with regret.

An experimenter may give his value of (μ/I) directly, or may give his value of g , which is (μ/I) expressed in terms of a peculiar unit. The peculiar unit is $eh/4\pi m_p c$, in which m_p stands for the rest-mass of the proton and the other symbols have their normal meanings. This unit got into the picture because there was a doctrine that (μ/I) for the proton ought to be just two of it. This was based on an analogy with the electron which, to the consternation of theorists and the complication of Nature, proved to be fallacious. Reported values of g range from nearly 6 to 0.143; the proton has one of the two highest values, the triton or nucleus of hydrogen 3 has the other. Since all these values may be described without much extravagance as being "of the order of 2," the use of g remains convenient.*

Many people say that they have measured μ . Formally this is all wrong, but practically it is usually all right, for in most if not all cases I is known from experiments of other kinds. Most of these people give the value of μ in "nuclear magnetons." This means that they are giving the value of gI , as is seen from the following equation which resumes in notation what I just said in words, and provides the definition of g :

$$\mu = gI(eh/4\pi m_p c) \quad (10)$$

The quantity in brackets is called "nuclear magneton."

Now that this tiresome but necessary passage is behind us, we can review the results.

Values of gI — or of some other of the quantities catalogued above — have been published for about forty nuclei. The values of gI available toward the end of 1950 were gathered together and published in an article to which I give the reference in a footnote.† The largest is about twenty-five times the smallest: this is a wide range of variation, yet not so wide as that of the nuclear charges or the nuclear masses. Isotopes of one another may have values nearly the same or considerably different; the same is true of isobars. Most of the values are positive: this

* It is perhaps not premature to mention that in optical spectroscopy and in electronic magnetic resonance, the symbol g is used with a similar but not an identical meaning.

† Pake, G. E., *American Journal of Physics*, **18**, pp. 438-52, pp. 473-86, 1950. The table is on p. 440 of the October issue.

means that the angular momentum and the magnetic moment are parallel. A few are negative: for these (one of which is the neutron) the angular momentum and the magnetic moment are anti-parallel.

These values of μ (for, I repeat, gI is μ expressed in terms of a particular unit) are useful as challenges and as aids to the nuclear theorists. They are challenges, because the μ -value of a given nucleus is something to be explained; they may be aids, because a theory may be fortified by giving the right value of μ or confuted by leading to a wrong one. Now, nuclear theory is difficult, and by and large it is not so far advanced that it can demand experimental values accurate to let us say, the legendary "sixth place of decimals." This is a piece of temporary good fortune, for two reasons.

First, the strength of the big field H may not be known with adequate accuracy at the place where the nuclei are. It can however be ignored if one is concerned only to measure the ratio of the (μ/I) values of two nuclei. The experimenter has then to put into his apparatus successively samples containing the two kinds of nuclei, or a single sample containing them both: the ratio of the frequencies at which the resonance-peaks appear is the ratio of the (μ/I) values, and H vanishes in the division. Often the comparison-nucleus is the proton, so that many published values of gI come ultimately from ratios in which gI for the proton stands in the denominator. Such ratios are frequently adequate for the testing of theories, and their accuracies may be very good indeed, even attaining the sixth significant figure. (The basic determination of μ for the proton itself will be mentioned in Part II.)

Second, the true field which the nuclei experience may be slightly different from the big field H , because of local fields within the substance. This is of course an admission that our fundamental equation, (5) or (9) in this article, can be wrong. So it can be, and this is a development that may be thought distressing. But such developments are almost the rule in physics, whenever the art of measurement is bettered; and in the present case the errors in equation (9) must be regarded as felicitous, for they lead to some of the most fascinating applications of nuclear resonance.

Thus when ammonium nitrate, NH_4NO_3 , is put into the apparatus, there are two peaks of nitrogen instead of one. They are not far apart — if for the frequency in use one is at $H = 10,000$ gauss the other is at 9,997. The formula NH_4NO_3 suggests, and the diagram of the molecule would confirm if we had it here, that the two nitrogen nuclei are differently placed in the molecule: one may say that they have different atomic surroundings. Thus the position of either of the peaks is dis-

tinctive not of nitrogen alone, but of nitrogen in its particular surroundings. These same surroundings might recur in several different types of molecule, or might be confined to one. The formula of ethyl alcohol may be written as $\text{CH}_3\cdot\text{CH}_2\cdot\text{OH}$. This compound presents three proton-peaks, Fig. 7, separated by a few per cent of one gauss when the big field is of the order of 10,000: they have been ascribed to protons in the three "groups" CH_3 and CH_2 and OH . To identify a group is to perform a process of chemical analysis, and this is a nascent application of nuclear magnetic resonance.

This is a good place to speak of the efficacy of nuclear resonance in revealing the presence of chemical elements or of individual isotopes. The proton is one of the easiest nuclei to discern in this way, largely owing to its relatively high magnetic moment. It has been calculated that $2\cdot 10^{16}$ protons suffice to give a detectable "signal" by the induc-

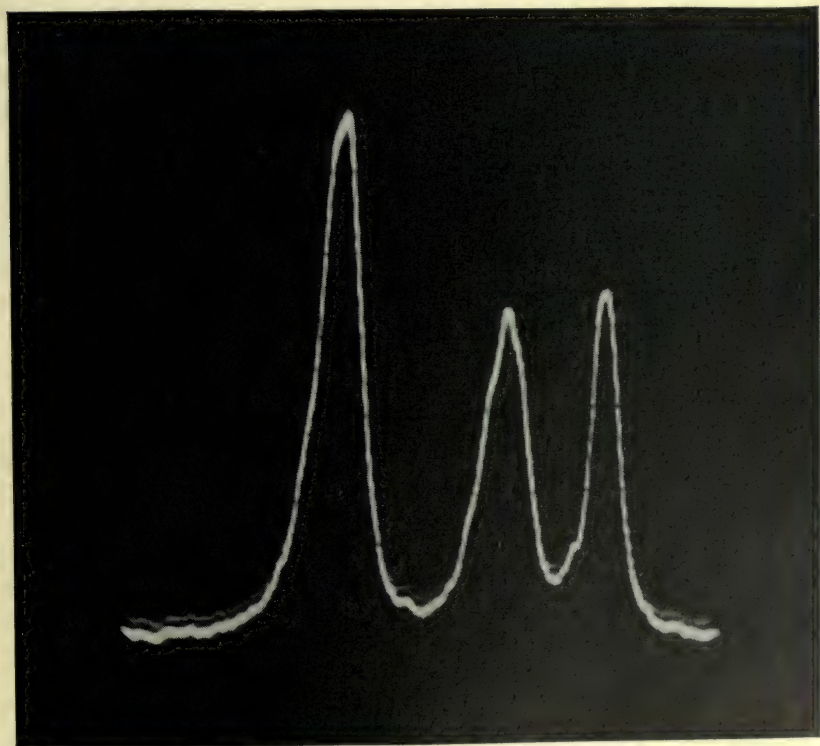


Fig. 7 — Breakup of the proton resonance peak of ethyl alcohol into three peaks, each believed to arise from protons in distinctive "groups" within the molecule. (Courtesy of M. E. Packard).

tion-method; a small capsule of gaseous hydrogen at a pressure of only one atmosphere will show the resonance of protons. The second isotope of hydrogen is normally present in that substance in an abundance of only 1.5 parts in ten thousand, the second isotope of oxygen exists only in an abundance of four parts in ten thousand; neither was discovered for more than a decade after the search for isotopes was well under way; but both of them have been detected by nuclear resonance.

Another application is to crystallography. In the crystal called gypsum, each proton is exposed to a magnetic field of the order of ten gauss from its neighboring protons. The resonance-peak is split into two or three or even four, depending on the inclination of the big field to the crystal axes. It would take many pages to describe this effect in detail, but it is so intelligible that one may deduce from it the positions of the protons in the crystal lattice. Nuclear resonance in fact seemed called to play a great role in crystallography, since the principal tool of the crystallographer has been the diffraction of X-rays, and this will not disclose the presence nor *a fortiori* the locations of protons in a crystal lattice. However this promising child of resonance has apparently been throttled in its cradle, for the still newer art of neutron-diffraction has proved itself adequate for finding the protons in a lattice.

Another application is to the measurement of magnetic field strengths. One sees that if proton-resonance is produced at a measured frequency in a steady field of which the magnitude H is unknown, H may be determined by equation (3) with an accuracy contingent on the accuracy with which μ_p is known, and this is pretty high. This has become a common method of measuring magnetic field strengths.

RELAXATION

If anyone were asked to guess the most important use of nuclear magnetic resonance, he would have two good reasons for choosing the study of relaxation. More pages of the scientific journals have been devoted to it than to any other application. Moreover, the discoverers spoke of it almost as soon as they spoke of the discovery; one has the feeling that they were so confident of the discovery, that as soon as it was made they considered it much less important for its own sake than as a tool.

"Relaxation" is a word that entered long ago into physics. Its general meaning is the gradual self-adjustment of a system to a sudden change in conditions. In the immediate instance the system is our sample in the big magnetic field; the sudden change in conditions is the starting or

the stopping of the oscillating field; and what gradually adjusts itself is the distribution of the protons between the up and the down orientations. Now I will describe an experiment such as has been performed on protons in water.

Let the sample be placed in the big field some time — several hours will always be ample — before the experiment is to begin. The experimenter should know in advance the frequency of the Larmor precession, so that he can apply the oscillating field of proper frequency as soon as he and the sample are ready. The sample then enters into what I will call “the state of resonance.” The experimenter is to measure the height of the peak as soon as the oscillating field is switched on: I call this initial stature A_0 . The big field and the alternating field are now *both* to be kept on. The height of the peak, A , is to be measured from time to time, say once every tenth of a second (this has been done with movie techniques). It is found that A is a declining function of time; the peak is shrinking.

After a while, let the alternating field be switched off while the big field continues to be on. The state of resonance is now suspended. Again the height of the peak is to be recorded every tenth of a second. Needless to say, the alternating field must be on while the record is being made, but it shall be off all of the rest of the time, which is most of the time. It will be found that the peak is growing again. It is, in fact, trending back to its initial stature A_0 , and the law of its rise is the exponential law:

$$A = A_0 [1 - \exp(-t/T_1)]^* \quad (11)$$

The constant T_1 , which this experiment determines, is called the “spin-lattice relaxation-time.” “Lattice” will be recognized as a term appropriate to crystals: in the literature of this subject it is however applied to all solids and liquids. Its meaning in this field may be put as follows: the “lattice” is all of the sample except the nuclear spins.

The actual experiment is not usually done quite as I just described it. The alternating field does not have to be switched on or off, because if its frequency is far from the Larmor frequency it is practically ineffectual. If the observer wants to end the state of resonance, he displaces H or ν away from the resonance-value; if he wants to restore it he brings H or ν back to the resonance-value. By modulating the big field with say a 60-cycle frequency, he may pass the system briefly

* This formula implies that $A = 0$ when $t = 0$; the reader can recast it to cover the general case in which $0 < A < A_0$ at $t = 0$.

through the state of resonance 60 times in a second; and it is possible to record and measure A on every such passage.

The fall and the rise of A are due to a cause so obvious that the reader has probably guessed it already. The peak, we recall, is a peak of absorption due to the turning of "up" protons into the "down" direction. When an up proton is turned into the down direction it goes out of business as an absorber, and continues out of business so long as it remains in the down direction. Since there is a fall and since there is a rise, the sojourn in the down direction must be neither zero nor infinite. If it were zero there would be no fall, and if it were infinite there would be no rise. The shrinkage of the peak in the presence of magnetic resonance, and the growth of the peak after resonance is discontinued, are signs that the sojourn of a proton in the down direction is finite but not zero. We divine already that T_1 is a measure of the average of this sojourn.

The foregoing may seem to imply that the height A of the peak is proportional to the number of up protons in the sample; but this is not so, and A is proportional to something else which I call the "margin." To present it I use N_u for the number of upward-pointing protons, N_d for the number of downward-pointing protons, N_0 for their constant sum, μ in the sense defined before, k for Boltzmann's constant; and I write down the fundamental theorem of Boltzmann:

$$N_u/N_d = \exp(2\mu H/kT) \quad (12)$$

The "margin" is $(N_u - N_d)$. We find:

$$\begin{aligned} N_u - N_d &= N_d [\exp(2\mu H/kT) - 1] \\ &= N_0 (\mu H/kT) \text{ approximately} \end{aligned} \quad (13)$$

We have approximated by supposing μH to be very small compared with kT , which it is indeed; and by supposing N_u and N_d each to be nearly half of N_0 —this second approximation is retroactively verified, for on substituting (for instance) 20,000 oersteds for H and room-temperature for T , one finds that out of two million protons selected at random a million plus seven are pointed up and a million minus seven are pointed down. The margin is thus 14 in two millions; but it may also be regarded as seven in two millions, since if seven protons out of two million should be turned down the margin would vanish and the peak would vanish with it.

Why is the stature of the peak proportional to the margin and not to N_u ? The point is, that in addition to turning protons from the up

direction to the down direction, the alternating field also helps protons to turn from the down to the up direction. Processes of the first kind involve absorption, as we already know; processes of the second kind involve release of energy. What the detector receives, and what the peak makes manifest, is the net of the absorptions over the releases. (This effect of the alternating field in helping protons from the down to the up direction is called "stimulated emission").

Now we must scrutinize equation (13) more closely. It is evident that T stands for an absolute temperature: the question is, what is it the temperature of?

One supposes perhaps that T is the temperature of the sample — that is to say, the temperature which would be shown by a thermometer stuck into the sample or possibly into a surrounding bath. And this is indeed what is supposed when the peak has the stature A_0 , signifying that the sample has stood long enough in the big field undisturbed by resonance or anything else. When A is A_0 and T is the temperature of the sample, (13) is right. But when A is less than A_0 because the peak is falling, has fallen or is rising, we must choose between saying that (13) is not right, and saying that (13) defines a temperature which is to be called the temperature of the spinning nuclei, or the "spin temperature" for short.

The second choice is made; and this is the most vivid language in which to describe the situation. In this language we say that the resonance elevates the spin-temperature, or heats up the spins; and that after resonance ceases, the spins cool down to the temperature of the lattice. Thus the study of relaxation becomes the study of the heating and the cooling of the spins with respect to the lattice — "lattice" being defined, I recall, as everything in the sample except the spins.

Recorded values of T_1 range from times of the order of hours down to times of the order of ten-thousandths of a second. The highest are exhibited by protons in ice at extremely low temperatures; protons in water have $T_1 = 2.33$ seconds; the lowest values are found in the presence of "magnetic impurities." The typical dependence of T_1 on temperature is represented by a curve with a single minimum.

The importance of "magnetic impurities" derives from the agent of relaxation. Relaxation is operated normally by the varying magnetic fields whereby the nuclei act on one another; these vary, as I shall presently say more fully, because the nuclei are wiggling in thermal agitation. But the magnetic fields of nuclei are comparatively small, and therefore normal relaxation is comparatively slow. Much bigger is the magnetic field of an electron, for the magnetic moment of an electron

is 660 times as great as that of a proton. Now, it is indeed true that all atoms and molecules contain electrons; and one may properly wonder why they do not always dominate the relaxation. The reason is that in most atoms and most molecules the electrons are paired "anti-parallel" so that their magnetic moments neutralize each other. (We are to see in Part II that this confines the electronic type of resonance to certain very specialized types of substances). One can however introduce into a substance, water for example, atoms or ions for which the neutralization is incomplete. These have much bigger magnetic moments and magnetic fields than any nucleus, and they speed up the spin-lattice relaxation. There is one special case which I treat in more detail, because of its relevance in this connection and its importance in solid-state physics.

There are crystals, of fluorite for instance, which occur colorless in Nature. These may be colored by exposing them to X-rays, or in other ways which we pass over here; and colored examples may also be found in Nature. Solid-state physicists have long been acquainted with these colorations, which they ascribe to what they call "F-centers." Various lines of reasoning have converged on the conclusion that an F-center is a cavity in the lattice (now I am using "lattice" in the normal sense, that of the crystallographers) in which a free electron is batting around like a wild animal in a cage. If this is so, then coloration of a colorless crystal by X-rays or otherwise should reduce its relaxation-time, and naturally-colored crystals should have lesser values of T_1 than those that are colorless. Experiment has ratified these inferences, and thus nuclear magnetic resonance has come to confirm the theory of the F-centres. So also has electronic resonance, since the F-centres display it with extreme clarity; but this is a topic for Part II.

The cause of relaxation has now been identified as the thermal agitation of the substance, working through the variation-in-time of the magnetic fields which act on every nucleus, weakly from its neighbor nuclei and strongly from any uncompensated electron that happens to be in the vicinity. In gases and liquids the nuclei cruise around, and so do the "magnetic impurities" if there are any; fieldstrengths change swiftly and relaxation tends to be rapid. In solids the atoms and their nuclei vibrate around fixed positions, and thermal agitation has come to be interpreted in the following way.

Nowadays one thinks of the solid sample as quivering with compressional waves, and perhaps torsional waves as well. These constitute the thermal agitation of the sample, and their various frequencies form its elastic spectrum. From this broad band of frequencies we isolate, in

mind, the one which agrees with the frequency of the Larmor Precession. Think now of any two adjacent protons. The distance r between them will fluctuate in a complicated way as time goes on, but in this complicated motion we distinguish, again in mind, the component which has the frequency of the Larmor Precession. Well, according to earlier theory it is this component of the vibrations — be they interpreted as vibrations of the whole lattice or of two neighboring protons relative to each other — which helps the protons to turn from down to up, or for that matter from up to down. This is the channel by which energy passes to and fro between the lattice and the spins.

On submitting this idea to calculation it was found to give values of T_1 that are far too long. The next recourse was to take into account the "beat tones." Choose any frequency whatever in the elastic spectrum, and then another frequency differing from the first one by the frequency of the Larmor Precession. The sum of the two will present a beat frequency equal to that of the precession. This is a mathematical statement which seems as empty of physical meaning as — well, as seemed in its turn the assertion that the alternating magnetic vector H_1 directed along the x -axis is the sum of two circularly-polarized vectors. But the force between two neighboring protons is not linear (it is proportional to the inverse third-power of the distance r), and this gives physical meaning to the statement: the two frequencies conjointly act as if the beat-frequency were present. When these channels of communication between the spins and the lattice are added to the one first thought of, the calculated relaxation times come down into the order of magnitude of the real ones. More in the way of precise agreement can scarcely be hoped for, because of the effect of impurities on T_1 .

Two other topics in the field of spin-lattice relaxation must at least be mentioned.

Some of the known values of T_1 are too low to be measured by observing the rise and fall of the resonance-peak. To indicate how these are measured, I recall that the energy of a wave-train is proportional to the square of its amplitude, H_1^2 in the present case. To speak of protons for simplicity: if T_1 were zero the number of protons in the "up" orientation would always be the same, and hence the height A of the peak would be proportional to H_1^2 . But since T_1 is not zero the number of protons capable of absorbing goes down as H_1^2 goes up; and the curve of A against H_1^2 starts off tangent to the ideal straight line for $T_1 = 0$, but is concave-downward, drops away from the straight line and eventually will cease to rise.

We may pursue the argument one step farther. Here is the equation

for the rate of change of N_u , the number of "up" protons:

$$dN_u/dt = (1/T_1)(N_{0u} - N_u) - bH_1^2 N_u \quad (14)$$

Here N_{0u} stands for the number of "up" protons in the condition of equilibrium between the spin-temperature and the lattice-temperature. If we were dealing with the rise of the peak after the alternating field is shut off, the second term on the right would vanish, and we should be back at equation (11). We are however dealing here with the fall of the peak when the alternating field is on. One sees that eventually N_u will reach a constant value — it is said to "saturate" — and the peak a constant stature. If this saturation-value is measured and the value of b is known, T_1 can be computed. The saturation-value is measured, and b is determined from quantum mechanics.

Though I have tried to avoid giving the impression that the stature of the peak necessarily has its equilibrium-value A_0 before the oscillating field is first applied, I may not have quite succeeded. It would be a miracle if A were equal to A_0 at the moment when the sample is first put into the big field. Time must be allowed for the nuclei to adjust themselves or "relax" to the big field: it was because of this that I said at the beginning that the sample was to be placed in the big field several hours (a generous allowance of time, by the way) before the application of the alternating field. One would expect in general to find A much smaller than A_0 , when the sample has just been exposed to the big field; on the other hand it could be greater than A_0 if the sample had previously been exposed to a field of greater strength than the field of the experiment.

Conceivably one might miss the peak altogether by looking for it too soon, if the relaxation-time were long; or by looking for it too late, after it had been reduced to its "saturation" stature. It may be that early attempts to find magnetic resonance were frustrated in these ways. Such dangers are now avoided by mixing the sample deliberately with magnetic impurities in order to diminish the value of T_1 : the peak of Figure 2 was obtained with water mixed with ferric nitrate. There is some reason also to conjecture that nuclear magnetic resonance might have been sought and found some years earlier than it was, but for an imperfect theory which indicated that the spin-lattice relaxation-time would be so long as to make it hopeless to look for the peak.

Students of the literature will find many allusions to another type of relaxation — the "spin-spin" relaxation, with a relaxation-time denoted by T_2 . Except for the bare statement that the breadth of the peak varies inversely as the spin-spin relaxation-time, this topic must be left for some other occasion. It may be mentioned here, even though not

explained, that the line-breadth diminishes suddenly when the substance melts, and may also decline at one or more temperatures where the substance is still a solid. Such temperatures are considered to be those at which some special type of molecular motion begins.

References and acknowledgements are to be appended to the second part of this article. Two names will however be mentioned in this place, those of Felix Bloch and Edward M. Purcell; for these are the names of the physicists to whom, on November 6, 1952, was awarded the Nobel Prize for the discovery of nuclear magnetic resonance by the techniques here respectively denoted as those of "nuclear induction" and "nuclear resonance absorption."

Delay Curves for Calls Served at Random

By JOHN RIORDAN

(Manuscript received March 11, 1952)

This paper presents curves and tables for the probability of delay of calls served by a simple trunk group with assignment of delayed calls to the trunks at random and with pure chance call input. These are contrasted with the classic results of Erlang ("Erlang C") which are based on service in order of arrival. Trunk holding times for both have an exponential distribution. The theoretical development for computation of the curves is directed to the determination of the moments, which seem to be a natural means of simplification.

1. INTRODUCTION

One of the classic results in the study of telephone traffic is the formula for delay given by the Danish engineer A. K. Erlang¹ in 1917. This is for random call input to a fully accessible simple trunk group with the trunk holding time exponential and calls served in the order of arrival. A proof for this formula and a set of curves for its use have been given by E. C. Molina.²

In many switching systems it is not feasible to fully realize this ethical ideal of first come, first served, and it has long been of interest to determine delays on another basis. The contrasting assumption is of calls picked at random, which is again an idealization but in large offices appears to be called for, as a bound for the service actually given.

The first attempt to formulate the last seems to be that of J. W. Mellor.³ While his basic formulation is incomplete, it offers a useful approximation to the complete results, particularly in the most interesting region of heavy traffic, and will be referred to here as the "Mellor approximation." A complete formulation due to E. Vulot⁴ appeared in 1946 and included both the fundamental differential recurrence relation and formulas for delay probabilities for small delays. For completeness, these are repeated below. F. Pollaczek⁵ has given a development of Vulot's work directed toward determining an asymptotic delay formula.

Considerable further theoretical work has been necessary to obtain the results given here. Vaultot's differential recurrence relation, which formulates the probability of delay at least t of a call which arrives when n other calls are waiting, has no simple solution. By approximate methods, it was possible to use a differential analyser to determine these probabilities for small values of n . But it was not feasible in this way to cover the whole range of interest, and these results were supplemented by approximations for large n , which are described below. Finally the delay for an arbitrary call was obtained by summing on n .*

These results are not reported here, because the attempt to verify the accuracy attained led to formulation of the moments of the delay curves and this in turn to the representation of the curves as sums of exponential curves, with great simplification of the calculations required. As will appear, two exponentials furnish a sufficient approximation except for heavy traffic.

2. DELAY CURVES

The delay distribution on calls delayed for occupancy levels (defined below) from 0.1 to 0.9 in steps of 0.1 is shown in Fig. 1. The abscissae are derived time units which seem to be natural to the problem: $u = ct/h$, with c the number of trunks and h the average holding time. The ordinates, on a logarithmic scale, are conditional probabilities that a call delayed will be delayed at least u , that is, values of a function $F(u)$; the logarithmic scale is chosen to emphasize the dominantly exponential character of the curves. The occupancy level α is the ratio a/c where a is the average call input in average holding time h .

Fig. 1 is a master curve for all eventualities and may be changed to working curves for various sizes of trunk groups. For the construction of these curves Table I, from which Fig. 1 was made, and which also compares present results with those for calls served in order of arrival, is convenient. A more elaborate table will be given later. For the convenience of the reader, it may be noticed that for order of arrival service $F(u) = e^{-u(1-\alpha)}$.

The striking feature of Table I is the increase in delay time for random service, which becomes more pronounced with decreasing $F(u)$ and increasing occupancy (or traffic) level, α . The increase throughout the table is an effect of the limitation to small values of $F(u)$. For given

* Thanks are due to George W. Abrams for directing this work, to Dr. Richard W. Hamming for transforming the equations into forms suitable for the differential analyser and for supervising its operation, and to Miss Catherine Lennon for a great deal of calculation.

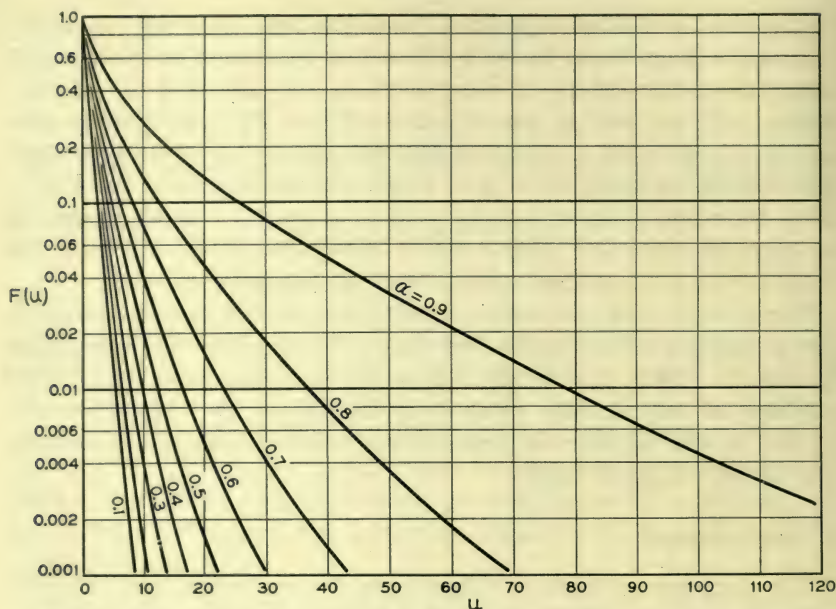


Fig. 1—Delay curves for random service. $F(u)$ = conditional probability of delay at least u ; $u = ct/h$, c = no. trunks, h = av. holding time, a = call input in time h , $\alpha = a/c$.

α , the delay curves for order of arrival and random service include the same area, which is in fact equal to the mean delay (of calls delayed), $(1 - \alpha)^{-1}$. Since $F(0) = 1$ for both, and the random service curve decreases more slowly for large u , the curves must intersect at some point, say for $u = u_0$; for $u < u_0$, the o.a. curve must be above the ran-

TABLE I — DELAY-TIME AND RANDOM SERVICE

Delay Times, u , for given $F(u)$ and α and for order of arrival (o.a.) and random service.

α	$F = 0.1$		$F = 0.01$		$F = 0.001$	
	o.a.	Random	o.a.	Random	o.a.	Random
0.1	2.56	2.58	5.12	5.47	7.68	8.60
0.2	2.88	2.91	5.76	6.57	8.63	10.68
0.3	3.29	3.34	6.58	8.05	9.87	13.35
0.4	3.84	3.91	7.68	10.04	11.52	16.95
0.5	4.61	4.68	9.21	12.89	13.82	22.09
0.6	5.76	5.82	11.51	17.25	17.27	29.97
0.7	7.68	8.28	15.36	23.14	23.03	43.33
0.8	11.51	12.57	23.03	36.80	34.54	70.29
0.9	23.03	25.99	46.05	77.24	69.08	156.63

dom curve. This is shown in Fig. 2 for $\alpha = 0.9$, but the logarithmic scale for $F(u)$ obscures the equality of area.

The character of the comparison may be clearer if the picture is changed. Consider a department store counter with c clerks (corresponding to c trunks) in attendance. The time for a sale corresponds to the trunk holding time, and the rate of arrival of customers is like that of call input. For service in order of arrival customers are given serially numbered tickets on arrival; for random service, these tickets may be supposed drawn from a hat, or numbered from a series of random numbers, or since aggressiveness and the clerks' attention are subject to devious rule, it may be that no attention at all to order of service is equivalent to random service.

The fact that the average delay is independent of the order of service may be explained roughly by saying that the average rate at which waiting lines are removed depends only on the average rate of arrival of customers and the rate at which they are served. Notice however that service at random causes more variable delays (the second and all higher moments are larger than for order of arrival service). Thus with random service the proportion of waiting customers receiving quick

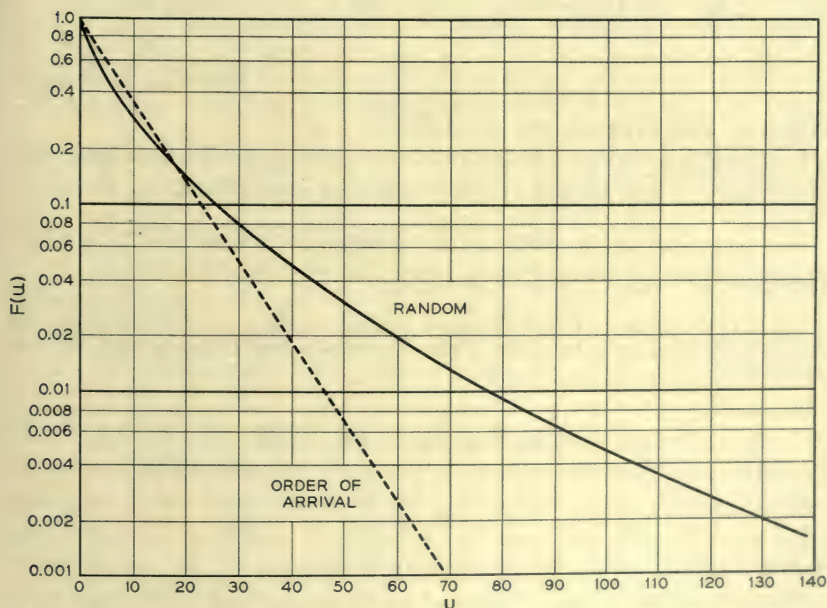


Fig. 2—Comparison of delay curves for order of arrival and random service; $\alpha = 0.9$.

service is increased (over order of arrival) but this is achieved at the cost of making other customers wait much longer.

Service in order of arrival has the advantage to the customer that his delay is independent of all who come after him, and this is particularly appreciated in times of heavy crowding when long delays are possible for random service. In Table I, these crowded conditions correspond to small values of $F(u)$ or large values of α , or both. In this picture it seems intuitively clear that much longer delays are possible for random service, for those unlucky customers who keep missing their turn. (Of course, a more realistic model would also include the effects of customers leaving before service, a factor of considerable telephone interest also.)

As noted at the start of this section, $F(u)$ is a conditional probability, the probability of delay at least u of a call that is surely delayed. To obtain unconditional probabilities of delay, $F(u)$ is multiplied by the probability that all trunks are busy, which is the probability that a call is delayed. This probability is given by a well-known formula due to Erlang and customarily written as

$$C(c, a) = \frac{a^c}{(c-1)! (c-a)} \left[1 + \frac{a}{1!} + \frac{a^2}{2!} + \cdots + \frac{a^{c-1}}{(c-1)!} + \frac{a^c}{(c-1)! (c-a)} \right]^{-1}$$

Tables of this function are available*.

Finally it may be noticed here that for random service and light traffic (roughly, α less than 0.7), with sufficient approximation

$$F(u) = \frac{1}{2}(y_1 e^{-u(1-\alpha)y_1} + y_2 e^{-u(1-\alpha)y_2})$$

with $y_1 = 1 - \sqrt{\alpha/2}$, $y_2 = 1 + \sqrt{\alpha/2}$.

* But there seems to be no extensive tabulation. However, the table for the Erlang B function made by Conny Palm (Stockholm, 1947) may be used with the relations

$$\begin{aligned} \frac{1}{C(c, a)} &= \frac{1}{B(c, a)} - \frac{1}{B(c-1, a)} \\ &= \frac{a}{c} + \frac{1 - (a/c)}{B(c, a)} \end{aligned}$$

Notice that $C(c, a)$ also has the recurrence relation

$$\frac{1}{C(c, a)} = \frac{-1}{c-1-a} + \frac{(c-a)(c-1)}{a(c-1-a)C(c-1, a)}$$

3. BASIC FORMULATION

As noted above, the following notation is used: c is the number of trunks, h is the average holding time (the distribution of holding times is exponential) and a is the average number of calls arriving in time interval h . Then, if $F_n(t)$ is the probability of delay at least t of a call arriving when n other calls are waiting, the differential recurrence relation given by Vaultot is

$$\frac{dF_n(t)}{dt} = \frac{n}{n+1} \frac{c}{h} F_{n-1}(t) - \frac{c+a}{h} F_n(t) + \frac{a}{h} F_{n+1}(t) \quad (1)$$

This may be derived as follows. Consider the interval dt after the epoch of arrival of the call in question. In this interval three events may occur: (i) a call may arrive, (ii) a trunk may be released, or (iii) neither of these. The probability of a call arrival is $(a/h)dt$ and if a call arrives the delay function is $F_{n+1}(t - dt)$. The probability of a trunk release, because of the assumption of exponential holding time, is $(c/h)dt$, and if a trunk is released the number of waiting calls is reduced by one; the probability that the call seizing the waiting trunk will not be the call in question is $n/(n+1)$. Finally the probability of the third event is $1 - (c+a)dt/h$. All this is summarized in the differential relation

$$F_n(t) = \frac{a}{h} dt F_{n+1}(t - dt) + \frac{n}{n+1} \frac{c}{h} dt F_{n-1}(t - dt) + \left(1 - \frac{c+a}{h} dt\right) F_n(t - dt)$$

Passing to the limit gives equation (1).

Using new variables: $u = ct/h$, $\alpha = a/c$, equation (1) may be written more simply as

$$\frac{dF_n(u)}{du} = \frac{n}{n+1} F_{n-1}(u) - (1 + \alpha)F_n(u) + \alpha F_{n+1}(u) \quad (1a)$$

This equation is a mixed differential-difference equation of the first order as a differential equation and of the second order as a difference equation; hence three boundary relations are required. For the differential part, it is clear that $F_n(0)$, which is the probability of some delay of the test call, is unity for all n in question, that is, for all integral non-negative n . Also $F_n(u) \equiv 0$ for all negative n , is an obvious necessity, and, since F_n is a distribution function $F_n(\infty) = 1$. Finally the third

condition may be stated as

$$\lim_{n \rightarrow \infty} F_n(u) = 1, \quad \text{all } u$$

The probability of delay at least u of an arbitrary call is the sum on n of the product of the probability that n calls are waiting when the call arrives and the probability, $F_n(u)$, that for this condition the call is delayed at least u . The first probability (for statistical equilibrium) is known to be

$$(1 - \alpha) C(c, a) \alpha^n$$

where $C(c, a)$, as stated above, is the probability that all trunks are busy; $(1 - \alpha)C(c, a)$ is the probability that all trunks are busy and no calls are waiting. Hence the probability in question, say $f(u)$, is given by

$$f(u) = (1 - \alpha)C(c, a) \sum_0^{\infty} \alpha^n F_n(u)$$

or by

$$f(u) = C(c, a)F(u)$$

if

$$F(u) = (1 - \alpha) \sum_0^{\infty} \alpha^n F_n(u) \quad (2)$$

$F(u)$, like $F_n(u)$, is then a conditional probability, the probability at least u of a delayed call. Notice that, consistent with this, $F(0) = 1$.

It is interesting to notice that Mellor's basic equation, which in present notation may be written as

$$\frac{dG_n(u)}{du} = -\frac{1}{n+1} G_n(u), \quad (3)$$

follows from (1) if first it is supposed that $F_{n-1}(u) = F_n(u) = F_{n+1}(u)$ and then, for clarity, G_n replaces F_n . Hence, as indicated by the third boundary condition, it may be expected to be useful for large values of n . Its solution is

$$G_n(u) = e^{-u/(n+1)} \quad (4)$$

A somewhat better approximation may be determined by the MacLaurin series obtained by repeated differentiation of (1a) and evaluation

at $u = 0$; this is as follows

$$F_n(u) \approx 1 - \frac{u}{n+1} + \frac{\alpha}{2} \frac{(u)^2}{(n+1)} - \frac{\alpha(2\alpha-1)}{3!} \frac{(u)^3}{(n+1)} + \frac{\alpha(2\alpha-1)(3\alpha-2)}{4!} \frac{(u)^4}{(n+1)} - \dots \quad (5)$$

But this is the same* as:

$$F_n(u) \approx [1 - (1 - \alpha)u/(n+1)]^{1/(1-\alpha)} \quad (5a)$$

As α approaches unity, (5a) approaches (4). Equation (5a) has been used, for large values of α , in the direct computations mentioned above.

It may also be noted that for $\alpha = 0$, equation (1a) has the solution (now writing $F_n(u, \alpha)$ for $F_n(u)$)

$$F_n(u, 0) = \phi(u, n) - \frac{u}{n+1} \phi(u, n-1) \quad (6)$$

where $\phi(u, n)$ is the Poisson sum

$$e^{-u} \left(1 + u + \frac{u^2}{2!} + \dots + \frac{u^n}{n!} \right)$$

Finally, for completeness, note that for small values of u , the MacLaurin series for $F(u)$ is

$$F(u) = 1 - u \frac{1-\alpha}{\alpha} \log \frac{1}{1-\alpha} + \frac{u^2}{2} (1-\alpha) \left[2 - \frac{1-\alpha}{\alpha} \log \frac{1}{1-\alpha} \right] - \frac{u^3}{6} (1-\alpha) \left[1 + 3\alpha - (1-\alpha) \log \frac{1}{1-\alpha} - \sum_1^{\infty} \frac{\alpha^n}{n^2} \right] \quad (7)$$

4. MOMENTS

The k 'th moment (about the origin) of the delay density function which is $-F'(u)$ ($F(u)$ itself is a distribution function is defined as

$$M_k = \int_0^\infty u^k [-F'(u)] du, \quad (8)$$

$$= k \int_0^\infty u^{k-1} F(u) du, \quad k > 0,$$

the last by integration by parts.

* G. W. Abrams is due credit for noticing this.

Following (2), this may also be written as

$$M_k = (1 - \alpha) \sum_0^{\infty} \alpha^n m_{n,k}, \quad (9)$$

with

$$\begin{aligned} m_{n,k} &= \int_0^{\infty} u^k [-F'_n(u)] du, \\ &= k \int_0^{\infty} u^{k-1} F_n(u) du, \quad k > 0. \end{aligned} \quad (10)$$

First, notice that

$$m_{n,0} = - \int_0^{\infty} F'_n(u) du = F_n(0) = 1;$$

hence

$$M_0 = (1 - \alpha) \sum_0^{\infty} \alpha^n = 1,$$

showing that $F(u)$ is properly normalized.

Next, by integrating both sides of (1a) with respect to u from 0 to ∞ , and using the second form of (10) (with $k = 1$)

$$-(n+1) = nm_{n-1,1} - (n+1)(1+\alpha)m_{n,1} + (n+1)\alpha m_{n+1,1} \quad (11)$$

In the same way, after first multiplying (1a) throughout by u^{k-1} , it is found that

$$\begin{aligned} -k(n+1)m_{n,k-1} \\ = nm_{n-1,k} - (n+1)(1+\alpha)m_{n,k} + (n+1)\alpha m_{n+1,k} \end{aligned} \quad (12)$$

Unfortunately, neither (11) nor any other instances of (12) have simple solutions; nevertheless they may be used to determine M_k .

Consider first the simplest case, M_1 . If (11) is multiplied throughout by α^n and summed on n , the result may be written

$$\begin{aligned} -L_{10} &= \alpha L_{11} - (1+\alpha)L_{11} + L_{11} - L_{01} \\ &= -L_{01} \end{aligned} \quad (13)$$

where for convenience in writing and of later notation

$$\begin{aligned} L_{01} &= \sum \alpha^n m_{n,1} = (1 - \alpha)^{-1} M_1 \\ L_{11} &= \sum (n+1) \alpha^n m_{n,1} \\ L_{10} &= \sum (n+1) \alpha^n = D \sum \alpha^{n+1} = (1 - \alpha)^{-2} \end{aligned}$$

and $D = d/d\alpha$. Hence

$$M_1 = (1 - \alpha)^{-1}$$

This is the mean delay of calls delayed and as mentioned above is the same as for service in order of arrival.

In the general case*, the following notation is convenient

$$L_{0k} = \sum \alpha^n m_{n,k} = (1 - \alpha)^{-1} M_k$$

$$L_{jk} = \sum (n+1)(n+2) \cdots (n+j) \alpha^n m_{n,k}$$

Using the relations

$$\begin{aligned} n(n+2) \cdots (n+j) &= n(n+1) \cdots (n+j-1) + (j-1)n(n+1) \cdots (n+j-2) \\ &+ \cdots + (j-1)n(n+1) \cdots (n+j-i-1) + \cdots \\ &+ (j-1)!n, \\ n(n+1) \cdots (n+j-1) &= (n+1)(n+2) \cdots (n+j) - j(n+1)(n+2) \cdots (n+j-1) \end{aligned}$$

with

$$(j-1)_i = (j-1)(j-2) \cdots (j-i),$$

the summing of (12) is found to result in

$$\begin{aligned} kL_{j,k-1} &= [j - (j-1)\alpha]L_{j-1,k} - \alpha[(j-1)_2L_{j-2,k} \\ &+ (j-1)_3L_{j-3,k} + \cdots + (j-1)_iL_{j-i,k} + \cdots + (j-1)!L_{1,k}] \end{aligned} \quad (14)$$

But this may be simplified by multiplying through by j and subtracting from the same equation with j replaced by $j+1$; the result is

$$(j+1-j\alpha)L_{jk} - j^2L_{j-1,k} = kL_{j+1,k-1} - jkL_{j,k-1} \quad (15)$$

Notice that for $j=0$, $k=1$, $L_{01} = L_{10}$, as in (13). Notice also that

$$\begin{aligned} \alpha^{j-1}L_{j0} &= \sum (n+1) \cdots (n+j) \alpha^{n+j-1} \\ &= D \sum (n+1) \cdots (n+j-1) \alpha^{n+j} \\ &= D(\alpha^j L_{j-1,0}) \end{aligned}$$

so that

$$L_{j0} = jL_{j-1,0} + \alpha DL_{j-1,0} = j!(1 - \alpha)^{-j-1}$$

* This procedure is the development of a suggestion made by S. O. Rice.

Then the ratio

$$\begin{aligned} L_{0k}(L_{k0})^{-1} &= (1 - \alpha)^{k+1} L_{0k}/k! \\ &= (1 - \alpha)^k M_k/k! = R_k \end{aligned} \quad (16)$$

is the ratio of these moments to those for order of arrival service; the last relation is a definition. In the same way the ratio

$$L_{jk}(L_{j+k,0})^{-1}$$

might be considered, but to avoid fractions the following somewhat odd change of variables seems convenient:

$$({}^{j+k}_k)g_k L_{jk} = p_{jk} L_{j+k,0} \quad (17)$$

where $g_0 = g_1 = 1$ and

$$\begin{aligned} g_{2k} &= (2 - \alpha)^{2k-1} (3 - 2\alpha)^{2k-3} \cdots (k + 1 - k\alpha) \\ g_{2k+1} &= (2 - \alpha)^{2k} (3 - 2\alpha)^{2k-2} \cdots (k + 1 - k\alpha)^2 \end{aligned}$$

Notice that

$$g_{2k+1}(g_{2k})^{-1} = g_{2k}(g_{2k-1})^{-1} = (2 - \alpha)(3 - 2\alpha) \cdots (k + 1 - k\alpha) = D_{k+1}$$

the last being a definition, again.

Since

$$(1 - \alpha)L_{k,0} = kL_{k-1,0}$$

it follows from (15) that

$$\begin{aligned} (j + 1 - j\alpha)p_{jk} - j(1 - \alpha)p_{j-1,k} \\ = (g_k/g_{k-1})[(j + 1)p_{j+1,k-1} - j(1 - \alpha)p_{j,k-1}] \end{aligned} \quad (18)$$

By taking differences of this equation and writing

$$\begin{aligned} q_{0k} &= p_{0k} \\ q_{1k} &= p_{1k} - p_{0k} = \Delta p_{0k} \\ q_{2k} &= p_{2k} - 2p_{1k} + p_{0k} = \Delta^2 p_{0k} \\ q_{jk} &= \Delta q_{j-1,k} = \Delta^j p_{0k} \end{aligned}$$

a somewhat simpler recurrence relation is found to be as follows

$$\begin{aligned} (j + 1 - j\alpha)q_{jk} &= (g_k/g_{k-1}) \\ [j\alpha q_{j-1,k-1} + (j + 1 + j\alpha)q_{j,k-1} + (j + 1)q_{j+1,k-1}] \end{aligned} \quad (19)$$

Since $p_{j0} = 1$, all j , $q_{00} = 1$, and $q_{j0} = 0$, $j \neq 0$. From these boundary conditions, it follows at once from (19) that

$$q_{jk} = 0, \quad j > k.$$

By comparison of (17) and (16)

$$g_k R_k = p_{0k} = q_{0k}$$

A short table of the q 's is as follows:

j/k	0	1	2	3
0	1	1	2	$2(2 + \alpha)$
1	0	$\alpha(2 - \alpha)^{-1}$	$4\alpha(2 - \alpha)^{-1}$	$2\alpha(18 - 5\alpha - 4\alpha^2)D_3^{-1}$
2	0	0	$2\alpha^2(3 - 2\alpha)^{-1}$	$2\alpha^2(18 - 7\alpha - 2\alpha^2)(3 - 2\alpha)^{-2}$
3	0	0	0	$6\alpha^3 D_3^2 D_4^{-1}$

Continuation of this leads to the values of R_k listed in Table II. Notice that for $\alpha = 1$, by (18)

$$\begin{aligned} p_{jk}(1) &= (j+1)p_{j+1,k-1}(1) \\ &= (j+1)(j+2)p_{j+2,k-2}(1) \\ &= (j+1)(j+2) \cdots (j+k) \end{aligned}$$

since $g_k = 1$ for $\alpha = 1$ and $p_{j0} = 1$, all j . From this

$$p_{0k}(1) = g_k(1)R_k(1) = R_k(1) = k!$$

On the other hand, for $\alpha = 0$, $q_{jk} = 0$, $j > 0$ and, by (19)

$$q_{0k}(0)/g_k(0) = q_{0,k-1}(0)/g_{k-1}(0)$$

so that

$$R_k(0) = R_{k-1}(0) = R_1(0) = 1$$

5. MELLOR APPROXIMATION

It is useful to have the moments of the distribution corresponding to the Mellor approximation, since they serve as a guide. Here, following equation (4)

$$F(u) = (1 - \alpha) \sum_{n=0}^{\infty} \alpha^n e^{-u(n+1)^{-1}} \quad (20)$$

and

$$\begin{aligned}\exp - xM &= \sum_0^{\infty} M_k (-x)^k / k! \\ &= (1 - \alpha) \int_0^{\infty} du e^{-xu} \sum_0^{\infty} (n+1)^{-1} \alpha^n e^{-u(n+1)^{-1}} \quad (21) \\ &= (1 - \alpha) \sum_0^{\infty} \alpha^n [1 + x(n+1)]^{-1}\end{aligned}$$

Hence

$$M_k = k!(1 - \alpha) \sum_0^{\infty} (n+1)^k \alpha^n \quad (22)$$

These moments are expressible in terms of polynomials associated with the distribution of permutations into classes according to the number of readings left to right necessary to find the elements in standard order.⁶ Indeed the ratio

$$r_k(\alpha) = M_k (1 - \alpha)^k / k!$$

has the recurrence relation

$$r_{k+1}(\alpha) = (k\alpha + 1)r_k(\alpha) + \alpha(1 - \alpha)r'_k(\alpha) \quad (23)$$

and the first few values are as follows

$$\begin{aligned}r_1 &= 1 & r_3 &= 1 + 4\alpha + \alpha^2 \\ r_2 &= 1 + \alpha & r_4 &= 1 + 11\alpha + 11\alpha^2 + \alpha^3 \\ r_5 &= 1 + 26\alpha + 66\alpha^2 + 26\alpha^3 + \alpha^4\end{aligned}$$

Notice that $r_k(0) = 1$, $r_k(1) = k!$, just as for the precise results.

6. EXPONENTIAL SUMS

The shape of the delay curves, from direct calculation, and also from Mellor's results, suggests representation in exponential sums. If

$$F(u) = A_1 e^{-(1-\alpha)u/x_1} + A_2 e^{-(1-\alpha)u/x_2} + \dots \quad (24)$$

then

$$M_k \frac{(1 - \alpha)^k}{k!} = A_1 x_1^k + A_2 x_2^k + \dots \quad (25)$$

by a simple calculation. For k exponentials, $2k$ moments (including

M_0) may be fitted exactly by solution of $2k$ equations of form (25), as will be shown.

The first approximation ($k = 1$) is the order of arrival curve, say

$$F_1(u) = e^{-(1-\alpha)u}$$

which has $A_1 = x_1 = 1$, $A_k = x_k = 0$, $k > 1$, and matches M_0 and M_1 .

The next approximation ($k = 2$) is determined from equations

$$A_1 + A_2 = 1$$

$$A_1x_1 + A_2x_2 = 1$$

$$A_1x_1^2 + A_2x_2^2 = R_2$$

$$A_1x_1^3 + A_2x_2^3 = R_3$$

Eliminating A_2 from successive pairs,

$$A_1(x_1 - x_2) = 1 - x_2$$

$$A_1x_1(x_1 - x_2) = R_2 - x_2$$

$$A_1x_1^2(x_1 - x_2) = R_3 - R_2x_2$$

Eliminating A_1 from these,

$$x_1 + x_2 - x_1x_2 = R_2 \quad (26)$$

$$(x_1 + x_2)R_2 - x_1x_2 = R_3$$

or, writing $a_1 = x_1 + x_2$, $a_2 = x_1x_2$, so that $x^2 - a_1x + a_2 = (x - x_1)(x - x_2)$

$$a_1 - a_2 = R_2 \quad (26a)$$

$$a_1R_2 - a_2 = R_3$$

From the first of the second set of equations, and from symmetry (or from $A_1 + A_2 = 1$)

$$A_1 = \frac{1 - x_2}{x_1 - x_2} \quad (27)$$

$$A_2 = \frac{1 - x_1}{x_2 - x_1}$$

Taking R_2 and R_3 from Table II, it turns out that

$$x_1^{-1} = 1 - \sqrt{\alpha/2} = 2A_1 \quad (28)$$

$$x_2^{-1} = 1 + \sqrt{\alpha/2} = 2A_2$$

TABLE II - MOMENT RATIOS, CALLS SERVED AT RANDOM

$$R_k(\alpha) = M_k(1 - \alpha)^k/k!$$

$$R_1 = 1$$

$$R_2 = \frac{2}{2 - \alpha}$$

$$R_3 = \frac{2(2 + \alpha)}{(2 - \alpha)^2}$$

$$R_4 = \frac{4(6 + 5\alpha - 4\alpha^2 - \alpha^3)}{(2 - \alpha)^3(3 - 2\alpha)}$$

$$R_5 = \frac{4(36 + 60\alpha - 59\alpha^2 - 24\alpha^3 + 15\alpha^4 + 2\alpha^5)}{(2 - \alpha)^4(3 - 2\alpha)^2}$$

$$R_6 = \frac{8f_6(\alpha)}{(2 - \alpha)^5(3 - 2\alpha)^3(4 - 3\alpha)}$$

$$R_7 = \frac{8f_7(\alpha)}{(2 - \alpha)^6(3 - 2\alpha)^4(4 - 3\alpha)^2}$$

$$f_6(\alpha) = 432 + 972\alpha - 2016\alpha^2 - 437\alpha^3 + 1790\alpha^4 - 528\alpha^5 - 196\alpha^6 + 67\alpha^7 + 6\alpha^8$$

$$f_7(\alpha) = 10368 + 34560\alpha - 89208\alpha^2 - 32772\alpha^3 + 177926\alpha^4 - 104287\alpha^5 - 29260\alpha^6 \\ + 43876\alpha^7 - 9158\alpha^8 - 2039\alpha^9 + 588\alpha^{10} + 36\alpha^{11}$$

and the second approximation is

$$2F_2(u) = (1 - \sqrt{\alpha/2}) e^{-u(1-\alpha)(1-\sqrt{\alpha/2})} \\ + (1 + \sqrt{\alpha/2}) e^{-u(1-\alpha)(1+\sqrt{\alpha/2})} \quad (29)$$

which turns out to be a good fit for α roughly less than 0.7. Curiously the corresponding Mellor approximation has a more complicated expression.

Following the same procedure for three exponentials, it turns out that the correspondent to the set of equations (26a) is

$$\begin{aligned} a_1 R_2 - a_2 + a_3 &= R_3 \\ a_1 R_3 - a_2 R_2 + a_3 &= R_4 \\ a_1 R_4 - a_2 R_3 + a_3 R_2 &= R_5 \end{aligned} \quad (30)$$

with $a_1 = x_1 + x_2 + x_3$, $a_2 = x_1 x_2 + x_1 x_3 + x_2 x_3$, $a_3 = x_1 x_2 x_3$, that is, the symmetric functions.

Using Table II for values of the R 's, it is found that

$$\begin{aligned} a_1 &= (18 - 7\alpha - 2\alpha^2)(2 - \alpha)^{-1}(3 - 2\alpha)^{-1} \\ a_2 &= 18 & (2 - \alpha)^{-1}(3 - 2\alpha)^{-1} \\ a_3 &= 6 & (2 - \alpha)^{-1}(3 - 2\alpha)^{-1} \end{aligned} \quad (31)$$

x_1, x_2 and x_3 are then the roots of the cubic equation

$$x^3 - a_1x^2 + a_2x - a_3 = 0$$

The coefficients $A_i, i = 1, 2, 3$ are determined from equations like

$$A_1 = \frac{R_2 - (x_2 + x_3) + x_2x_3}{(x_1 - x_2)(x_1 - x_3)} \quad (32)$$

For the fourth approximation, matching 8 moments, the equations for the symmetric functions are

$$\begin{aligned} a_1R_3 - a_2R_2 + a_3 - a_4 &= R_4 \\ a_1R_4 - a_2R_3 + a_3R_2 - a_4 &= R_5 \\ a_1R_5 - a_2R_4 + a_3R_3 - a_4R_2 &= R_6 \\ a_1R_6 - a_2R_5 + a_3R_4 - a_4R_3 &= R_7 \end{aligned} \quad (33)$$

and x_1, x_2, x_3 and x_4 are roots of the quartic equation

$$x^4 - a_1x^3 + a_2x^2 - a_3x + a_4 = 0$$

Coefficients A_i are determined from equations like

$$A_1 = \frac{R_3 - (x_2 + x_3 + x_4)R_2 + (x_2x_3 + x_2x_4 + x_3x_4) - x_2x_3x_4}{(x_1 - x_2)(x_1 - x_3)(x_1 - x_4)} \quad (34)$$

It may be noted that

$$\begin{aligned} x_2 + x_3 + x_4 &= a_1 - x_1 \\ x_2x_3 + x_2x_4 + x_3x_4 &= a_2 - x_1(a_1 - x_1) \\ x_2x_3x_4 &= a_3 - x_1[a_2 - x_1(a_1 - x_1)] = a_4x_1^{-1} \end{aligned}$$

which gives the general structure.

It is worth noting that equations (33) may be used to determine the R 's if the a 's may be determined otherwise. As a matter of fact, they have led to the determination of R_6 and R_7 in the following way. The

results for $k = 2$ and 3 suggest that

$$a_4 = 4!(2 - \alpha)^{-1}(3 - 2\alpha)^{-1}(4 - 3\alpha)^{-1}$$

$$a_3 = 4a_4$$

Then by the first two of equations (33)

$$a_1R_3 - a_2R_2 = R_4 - a_3 + a_4$$

$$a_1R_4 - a_2R_3 = R_5 - a_3R_2 + a_4$$

the solutions of which are

$$a_1 = 4(24 - 23\alpha + 3\alpha^3) \quad [(2 - \alpha)(3 - 2\alpha)(4 - 3\alpha)]^{-1}$$

$$a_2 = 2(72 - 23\alpha - 10\alpha^2 - 3\alpha^3)[(2 - \alpha)(3 - 2\alpha)(4 - 3\alpha)]^{-1}$$

By the last two of equations (33), R_6 and R_7 are determined to be the values given in Table II, which have been verified independently. Note that for $\alpha = 0$, both R_6 and R_7 are 1, and for $\alpha = 1$, $R_6 = 6!$, $R_7 = 7!$

Table III tabulates, for $k = 2$ to 5, for convenience in avoiding fractions the symmetric functions b_{kj} related to those above by

$$b_{kj} = D_k a_j$$

with, as before,

$$D_k = (2 - \alpha)(3 - 2\alpha) \cdots [k - (k - 1)\alpha]$$

and $a_0 = 1$. The functions for $k = 5$ were obtained by a process like

TABLE III — SYMMETRIC FUNCTIONS FOR EXPONENTIAL
SUMS OF CALLS SERVED AT RANDOM

$k = 2$	$b_{20} = 2 - \alpha$ $b_{21} = 4$ $b_{22} = 2$	$k = 3$	$b_{30} = 6 - 7\alpha + 2\alpha^2$ $b_{31} = 18 - 7\alpha - 2\alpha^2$ $b_{32} = 18$ $b_{33} = 6$
$k = 4$	$b_{40} = 24 - 46\alpha + 29\alpha^2 - 6\alpha^3$ $b_{41} = 96 - 92\alpha + 12\alpha^3$ $b_{42} = 144 - 46\alpha - 20\alpha^2 - 6\alpha^3$ $b_{43} = 96$ $b_{44} = 24$		
$k = 5$	$b_{50} = 120 - 326\alpha + 329\alpha^2 - 146\alpha^3 + 24\alpha^4$ $b_{51} = 600 - 978\alpha + 329\alpha^2 + 146\alpha^3 - 72\alpha^4$ $b_{52} = 1200 - 978\alpha - 172\alpha^2 + 78\alpha^3 + 72\alpha^4$ $b_{53} = 1200 - 326\alpha - 172\alpha^2 - 78\alpha^3 - 24\alpha^4$ $b_{54} = 600$ $b_{55} = 120$		

TABLE IV — SYMMETRIC FUNCTIONS FOR EXPONENTIAL SUMS, MELLOR APPROXIMATION

$k = 2$	$a_1 = 3 + \alpha$ $a_2 = 2$	$k = 3$	$a_1 = 6 + 3\alpha$ $a_2 = 11 + 5\alpha + 2\alpha^2$ $a_3 = 6$
$k = 4$	$a_1 = 10 + 6\alpha$ $a_2 = 35 + 26\alpha + 11\alpha^2$ $a_3 = 50 + 26\alpha + 14\alpha^2 + 6\alpha^3$ $a_4 = 24$		
$k = 5$	$a_1 = 15 + 10\alpha$ $a_2 = 85 + 80\alpha + 35\alpha^2$ $a_3 = 225 + 200\alpha + 125\alpha^2 + 50\alpha^3$ $a_4 = 274 + 154\alpha + 94\alpha^2 + 54\alpha^3 + 24\alpha^4$ $a_5 = 120$		
$k = 6$	$a_1 = 21 + 15\alpha$ $a_2 = 175 + 190\alpha + 85\alpha^2$ $a_3 = 735 + 855\alpha + 585\alpha^2 + 225\alpha^3$ $a_4 = 1624 + 1604\alpha + 1194\alpha^2 + 704\alpha^3 + 274\alpha^4$ $a_5 = 1764 + 1044\alpha + 684\alpha^2 + 444\alpha^3 + 264\alpha^4 + 120\alpha^5$ $a_6 = 720$		

that sketched above, and without determining R_8 and R_9 . Notice that

$$\begin{aligned}
 b_{kj} &= k! \binom{k}{j}, & \alpha &= 0 \\
 &= j! \binom{k}{j}^2, & \alpha &= 1
 \end{aligned}$$

which may be proved independently. All values in Table III satisfy the recurrence relation

$$\begin{aligned}
 b_{kj} &= [k - (k - 1)\alpha]b_{k-1,j} + [k + (k - 1)\alpha]b_{k-1,j-1} \\
 &\quad - (k - 1)^2\alpha b_{k-2,j-2}
 \end{aligned} \tag{35}$$

which also satisfies the boundary relations for $\alpha = 0$ and 1 given above for all values of k .

The corresponding symmetric functions for the Mellor approximation are given in Table IV. These have the recurrence relation

$$a_{kj} = a_{k-1,j} + [k + (k - 1)\alpha]a_{k-1,j-1} - (k - 1)^2\alpha a_{k-2,j-2} \tag{36}$$

For $\alpha = 0$, the values are the signless Stirling numbers of the first kind, that is, the numbers given by the expansion of

$$(1 + x)(1 + 2x) \cdots (1 + kx).$$

For $\alpha = 1$, the results are the same as for the exact case, as given above.

TABLE V — APPROXIMATIONS TO DELAY FUNCTION $F(u)$ FOR RANDOM SERVICE

α	v							
	1	2	4	6	8	10	12	14
Two Exponentials								
0.1	.3590	.1351	.0220	.0041	.0008	.0002		
0.2	.3490	.1344	.0256	.0058	.0015	.0004	.0001	
0.3	.3392	.1332	.0291	.0079	.0023	.0007	.0002	.0001
0.4	.3292	.1315	.0325	.0101	.0033	.0011	.0004	.0002
0.5	.3190	.1293	.0357	.0125	.0046	.0017	.0006	.0003
0.6	.3085	.1265	.0386	.0151	.0061	.0025	.0010	.0004
0.7	.2978	.1232	.0412	.0177	.0078	.0035	.0015	.0007
0.8	.2868	.1193	.0434	.0203	.0097	.0047	.0022	.0011
0.9	.2756	.1148	.0451	.0229	.0118	.0061	.0031	.0016
Three Exponentials								
0.1	.3586	.1354	.0219	.0040	.0008	.0002		
0.2	.3491	.1356	.0254	.0057	.0014	.0004	.0001	
0.3	.3393	.1358	.0288	.0074	.0022	.0007	.0002	.0001
0.4	.3291	.1360	.0322	.0092	.0030	.0011	.0004	.0002
0.5	.3186	.1363	.0358	.0112	.0040	.0016	.0007	.0003
0.6	.3071	.1359	.0392	.0133	.0050	.0022	.0010	.0005
0.7	.2951	.1354	.0428	.0156	.0063	.0028	.0014	.0007
0.8	.2822	.1344	.0466	.0181	.0077	.0036	.0018	.0010
0.9	.2683	.1325	.0504	.0210	.0094	.0045	.0023	.0013

7. NUMERICAL RESULTS

Table V gives both two-exponential and three-exponential 4 decimal approximations to the delay function $F(u)$ for

$$\alpha = 0.1(0.1)0.9(0.1 \text{ to } 0.9 \text{ in steps of } 0.1)$$

and for

$$u(1 - \alpha) = 1(1)2(2)14,$$

in the same abbreviated notation.* The variable $v = u(1 - \alpha)$ is introduced to reduce the spread of these tables. It will be noticed that, as expected, the two orders of approximation agree closely for small values of α ; indeed, only for the three largest values of α are the differences appreciable from the engineering standpoint.

* The results for two exponentials, some of those for three-exponentials, and all special results given below, have been obtained by Miss Marian Darville, whom I also thank for her careful drawing of the curves. The entire three-exponential table has been computed independently by Miss Lennon.

For $\alpha = 0.9$, results for four exponentials have also been obtained and compare with those of Table V as follows (k = number of exponentials):

k	v							
	1	2	4	6	8	10	12	14
2	.2756	.1148	.0451	.0229	.0118	.0061	.0037	.0016
3	.2683	.1325	.0504	.0210	.0094	.0045	.0023	.0013
4	.2748	.1402	.0483	.0195	.0091	.0047	.0026	.0015

It is somewhat surprising that two exponentials should do as well as they do for large values of v (in fact for $v = 12$ and 14 better than three); a similar behavior appears in the following comparison of approximations on the Mellor basis, again for $\alpha = 0.9$

k	v							
	1	2	4	6	8	10	12	14
2	.2725	.1115	.0446	.0237	.0129	.0070	.0038	.0021
4	.2671	.1379	.0502	.0207	.0097	.0051	.0029	.0018
6	.2777	.1408	.0477	.0205	.0102	.0054	.0031	.0018

From these comparisons, it appears a relatively small number of exponentials is sufficient for engineering purposes. The curves of Fig. 1 are those for three exponentials, for uniformity.

BIBLIOGRAPHY

1. Erlang, A. K., Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler. *Elektrotekniker* **13**, p. 5, 1917; *The Life and Works of A. K. Erlang*. Copenhagen, pp. 138-155, 1948.
2. Molina, E. C., Application of the Theory of Probabilities to Telephone Trunking Problems, *Bell System Tech. J.*, **6**, pp. 461-494, 1927.
3. Mellor, J. W., Delayed Call Formulae when Calls Are Served in a Random Order. *P.O.E.E.J.* **25**, pp. 53-56, 1942.
4. Vaulot, E., Delais d'attente des appels téléphoniques traités au hazard. *Comptes Rend. Acad. Sci. Paris* **222**, pp. 268-269, 1946.
5. Pollaczek, F., La loi d'attente des appels téléphoniques, *Comptes Rend. Acad. Sci. Paris* **222**, pp. 353-355, 1946.
6. Riordan, J., Triangular permutation numbers. *Am. Math. Soc., Proc.*, **2**, pp. 429-432, 1951.

The Evaluation of Wood Preservatives Part I

Interpretation and Correlation of the Results of Laboratory Soil-Block Tests and Outdoor Test Plot Experience, with Special Reference to Oil-Type Materials

By REGINALD H. COLLEY

(Manuscript received September 22, 1952)

This paper offers a review and interpretation of laboratory and field experiments aimed at determining the necessary protective threshold quantities of wood preservatives. It details the procedure followed in the soil-block tests at the Bell Telephone Laboratories, Incorporated. Discussion of specific criticisms of the techniques involved and replies to these criticisms are included. The paper also presents for the first time a correlation of the results obtained from soil-block culture tests, outdoor exposure tests on stakes and on pole-diameter posts as well as pole line experience. It demonstrates that the same levels for toxicity-permanence requirements (thresholds) are obtained from the three different types of accelerated experimental evaluations. There is every reason to believe that the same limits apply for the outer inch of sapwood in pine poles in line.

TABLE OF CONTENTS

Introduction.....	121
A Short History of the Development of Laboratory Evaluation Procedures.....	124
Evaluation by Soil-Block Tests.....	132
General Procedures.....	132
Inoculation and Incubation Rooms.....	134
Soil Characteristics and Moisture Content.....	134
Even-Aged Cultures.....	138
Standard Test Organisms.....	138
The Scope of the Soil-Block Evaluation Test.....	139
Preparation of the Test Blocks—Manufacture.....	140
Test Block Selection for Density.....	141
Average Block Volume.....	141
Treatment of the Test Blocks.....	142
Retention Gradients.....	143
The Amount of Preservative in the Blocks.....	144

Block Density and Preservative Absorption.....	146
Weathering.....	150
Conditioning.....	152
Sterilization.....	152
Flow Chart for the Bioassay Test.....	152
Some Madison Test Results.....	154
Check Tests at the Murray Hill Laboratories.....	159
Across the Threshold.....	160
The Significance of the Results of Laboratory Soil-Block Tests on Oil- Type Preservatives.....	160
Bibliography.....	163

SUBJECTS TO BE COVERED IN PART II

Evaluation by Treated $\frac{3}{4}$ Inch Southern Pine Sapwood Stakes in Test Plots
Rating the Condition of the Stakes
Depreciation Curves for $\frac{3}{4}$ Inch Stakes
Estimating Threshold Retentions and Average Life
Evaluation by Treated Pole-Diameter Posts in Test Plots
Evaluation by Pole Test Lines and by Line Experience; Service Tests
Discussion
Density and Growth Rate
Size and Shape of the Test Blocks
Toluene as a Diluent for Creosote Treating Solutions
The Distribution of the Preservative in the Block
Heat Sterilization of the Treated Blocks
The Weathering of Creosote and Creosoted Wood
General Considerations; Creosote Fractions
Creosote Losses
Creosote Losses from Treated Blocks
Creosote Losses from Impregnated Filter Paper
An Interpretation of Creosote Losses
The Gross Characteristics of the Residual Creosotes in Soil-Block Tests of Weathered Blocks
The Evaluation of Greensalt
The Evaluation of Pentachlorophenol
Swedish Creosote Evaluation Tests
Shortening the Bioassay Test
Toughness or Impact Tests for Determining Preservative Effectiveness
Other Accelerated Bioassay Tests
Other Observations
Conclusions
Acknowledgments

INTRODUCTION

In discussing the problems involved in the evaluation of wood preservatives over the years, it has generally been found necessary to orient the audience — in this case the readers of this JOURNAL — in the field of biology, and particularly in the field of biological tests involving wood-

destroying fungi. It is impractical to expect from such tests the degree of accuracy in results that one would look for as a matter of course in certain types of well conducted physical or chemical experiments. One can, however, look for high reliability in the biological sense. In the half science, half art of wood preservation there is as yet no generally acceptable laboratory technique for measuring the preservative value of a given material. Although much development work has been done, both here and abroad, in an effort to promote standard laboratory procedures, their proponents have had very little success in bringing into line the techniques used in the various areas. The interest of the Bell System in establishing a standard bioassay test will become convincingly evident as this story unfolds.

When the first American telephone lines were built there was an adequate supply of naturally durable pole timber in northern cedar and chestnut forests. The chestnut trees have been killed by a fungus disease, the chestnut blight; and the chestnut supply failed completely about twenty years ago. Northern cedar trees are not straight enough nor large enough nor plentiful enough to meet the demands of the power and communication utilities, but they are still used to some extent in the Lake States area. Usually they are incised at the ground line by toothed machines; and they are then given a preservative treatment with creosote or with pentachlorophenol in petroleum to prolong the life of the butt and ground line section.

In the northern and western states the increasing demands for poles 35 feet and longer brought in western red cedar, a straight and nearly perfectly shaped pole tree. The present Bell System use of the species is relatively small, about 4 per cent of the total annual production. Butt treatment of western red as well as northern cedar began in earnest about thirty years ago. This procedure protects the ground section. Many western cedars are now full length treated because, although the species is durable, the tops and sapwood layers are subject to infection and decay, sometimes after a relatively short service life.

In the South and Southeast the great favorite is naturally the southern pine pole, full length pressure-treated with creosote. Such poles made their way in the Bell System as far north as Memphis and Washington by the turn of the century. Their use increased rapidly after World War I, and they moved into virtually all parts of the country. They now make up about 73 per cent of the telephone pole plant. New treatment procedures for southern pine employing pentachlorophenol in petroleum applied by pressure processes are now under way at a number of plants.

Pressure-treated Douglas fir and butt-treated western red cedar dom-

inate other species on the West Coast and in the Pacific Northwest, while pressure and non-pressure treated lodgepole pine poles are favored in the Mountain States area. Pressure-treated jack pine and ponderosa pine move into telephone plant in small quantities in the Lake States and in the California areas, respectively.

To render telephone service the Bell System has some 20,000,000 wood poles carrying its wires and cables. Many of these poles are used jointly with the power companies. Since poles of the joint use sizes are not available in sufficient quantities in the southern pine forest to meet all the demands of the utilities all of the time it is inevitable that western cedar, Douglas fir, lodgepole pine, red pine and western larch should move into various parts of the System, either for the direct and sole use of the Operating Companies or for joint use.

The pole plant is continually changing. Pole species from the Northwest vary greatly in their treatability and they are generally harder to treat than southern pine. It is not possible to use traditional creosote pressure treatments for some of these species without running the risk of objectionable exudation, or bleeding, of the creosote.

The development of practical specifications for the application of new preservatives such as pentachlorophenol and greensalt, as well as the various types of creosote, to all of the pole species now used in Bell System plant calls for setting as exactly as possible necessary protective quantities of the various preservative materials. This is particularly true in view of the fact that for normal telephone use as well as for joint use it is absolutely essential to deliver to the Operating Companies poles that are clean and satisfactory for use in all types of telephone lines, without compromising on the question of adequate physical life for the treated units. This purpose is back of the Laboratories' efforts to develop bioassay tests that come as close as practicable to measuring the necessary protective amount of any given preservative, and to predicting its relative permanence in poles and crossarms in plant.

It has been pointed out in earlier papers^{30, 31, 76} that Bell Laboratories' concept of preservative evaluation involves (a) laboratory evaluation tests, (b) test plot experiments with small stakes, (c) similar tests of pole size specimens, and (d) test lines selected for long time observation. The latter are chosen with the cooperation of the Operating Companies. Lumsden⁷⁶ has recently presented a summary of a quarter century of experience with pole-diameter posts in one test plot located at Gulfport, Mississippi. The principal aims of the present paper are to interpret the results of various laboratory methods of preservative evaluation, and to

indicate how these results may possibly be correlated with test plot and field experience.

A SHORT HISTORY OF THE DEVELOPMENT OF LABORATORY EVALUATION PROCEDURES

The practice of laboratory evaluation of wood preservatives developed along different lines in Europe and in the United States. Here the Petri dish method was the early favorite.^{61, 100} The basic scheme of this test is to use agar culture media containing gradient concentrations of the preservative material to be tested, and to employ various easily grown test fungi as indicators of inhibiting or lethal doses. The same scheme is employed with stoppered Erlenmeyer flasks. The fungus now known as Madison 517, formerly referred to as *Fomes annosus*, has been used most frequently as the standard test organism although other fungi were also used.¹⁰¹

In the culture phase of the European standard agar-block method³³ the test fungi are grown in Kolle flasks on a malt agar medium. The impregnated wood test blocks are supported on glass "benches" just above the surface of the agar and the growing test fungus. Wood pulp or paper boards saturated with malt extract are used by some investigators^{9, 54 (1)} in place of the agar medium alone. Generally an untreated block and a treated block are placed together in the same flask.

The concept of a test for wood preservatives that motivated the proponents of the German agar-block method was broad enough to include selection of the test blocks and test fungi, treatment and handling procedures except weathering tests, culture technique, determination of the protection boundary, and directions for reporting the results. Differences in the behavior of water solutions of single chemical compounds such as sodium fluoride, and of volatile oily preservatives such as the creosotes were recognized; and provision was made for dealing with both types of materials.

The formalizing of both the Petri dish agar method in the United States and of the agar-block method in Europe developed as a result of conferences called by Dr. Hermann von Schrenk, the first in St. Louis in 1929, and the second in Berlin in 1930. The Laboratories' representatives at the St. Louis conference were the writer and R. E. Waterman. The action taken at St. Louis was published by Schmitz in 1930.¹⁰⁰

In a previous paper⁹⁹ about a year earlier, Schmitz had discussed various laboratory test procedures, and had offered an "improvement" in the Petri dish technique based on the idea of preventing evaporation of volatile materials. Some of his statements at that time now seem by

hindsight to have something of the character of a judgment before trial; but their bearing on the questions under discussion and their possible effect in retarding the development of more realistic methods appear to be important enough to warrant quoting at this time. For example, with special reference to Petri dish agar tests he says:

"The determination of the toxicity of relatively volatile substances, such as coal tar creosote, is particularly difficult, owing to the control of the loss of preservative during the sterilization process. In order to prevent this loss, it is proposed to place the preservative in small sealed glass ampules, which are later broken to liberate the preservative to form preservative-agar mixtures of any desired concentration."

He considers laboratory tests of toxicity of preservatives to have little or no application in commercial practice, and his opinions are definitely stated as follows:

"Toxicity studies deal only with the poisonous properties of a wood preservative, and therefore they do not give a complete picture of the value of any particular substance as a wood preservative. . . .

"For commercial work, however, it is of interest to know the amount of material that must be initially injected into the wood to maintain the desired amount of preservative for a definite period of time. (Author's italics). Laboratory studies of the toxicity of wood preservatives do not give this information. Attempts to calculate the amount of material which must be injected into the wood from laboratory studies of toxicity are, therefore, based upon an erroneous conception of the value of such studies."

Writing about the laboratory use of impregnated blocks of wood in testing wood preservatives, which was already well under way in Europe, he says that by using wood one may obtain conditions more or less closely resembling but not identical with conditions in actual service; but one would not only have to use a solvent in treating to low retentions, but there would be difficulties in obtaining an even distribution of the preservative in the wood. Furthermore:

"Getting rid of the solvent would require considerable time, during which a considerable loss of creosote would occur. . . .

"The composition of the creosote in the impregnated wood after the solvent has evaporated may be quite different from that of the original sample. More important still, the movement of the solvent in the wood during drying would cause an uneven distribution of the creosote."

The reader will bear in mind that these opinions were expressed in advance of the St. Louis and Berlin conferences on laboratory evaluation methods. Schmitz repeated them essentially in his 1930 paper, saying, for instance:

"Toxicometric values are not in themselves an index of the wood preserv-

ing value of the substance tested. Other factors, such as leaching, volatility, chemical stability, penetrability, cost, cleanliness, etc., must all be considered in the final evaluation of a wood preservative."

With respect to the European wood block test, he felt that

"... until more confidence can be placed in the even distribution of the preservative in the test block(s) their use will be greatly limited."

He has maintained his arguments with a high degree of consistency in later papers, and they have unquestionably influenced American thought on laboratory procedures and their practical application.

The Petri dish method adopted as a possible American standard procedure at the 1929 St. Louis meeting followed closely the techniques that had been developed and published by Humphrey et al.,⁶¹ Batemen⁸ and Richards.⁹⁴ Bell Telephone Laboratories made an intensive study of the Petri dish method during this period. The data obtained were never organized for publication since it was felt that the required evaluation of toxicity and permanence of toxicity of preservatives could not be obtained by the Petri dish test.

European workers would accept neither the Petri dish test method nor Madison 517 as the test fungus. In 1931, about a year after the Berlin conference, and four years before Liese et al.⁷¹ reported on the task force development of the agar-block method, A. Rabanus of the I. G. Farbenindustrie Aktiengesellschaft, Germany, published his "Die toximetrische Prüfung von Holzkonservierungsmitteln" (Toximetric testing of wood preservatives).⁸⁶ A somewhat expurgated and amended translation of this paper was presented to the American Wood-Preservers' Association in 1933. In the writer's opinion much of the force of the Rabanus argument was lost in the translation. The emphasis on the relative merits of the agar toximetric test and of the agar-block test was considerably diluted; and the cautiously guarded but nonetheless positive philosophy on the possibilities of using the results of agar-block tests in actual wood preserving practice was made water thin.

Apparently there was an understanding that subsequent to the 1930 conference in Berlin⁷¹ tests by the agar-block method would be run in the United States. For this project samples of creosotes as well as Scotch pine wood blocks were sent to a number of workers; but to the writer's knowledge no treated blocks were ever tested, or if they were no results were ever published. At Bell Telephone Laboratories some of the untreated Scotch pine blocks were put through preliminary trials with the Kolle flask technique,¹²⁵ and also a considerable number of plate and flask agar toxicity tests were run with the two sample creosotes. The inconsistency of the results — as far as translation to practical wood

preservation was concerned — was a strong stimulant toward the Laboratories' development of a block test, referred to later. It was more or less general information at the time that agar toximetric tests with native and European strains of test fungi were being run in other laboratories in this country; but again — as far as the writer knows — the results were not published.

In the meantime, and, as in the case of the Rabanus article cited above, before the publication of the Liese⁷¹ report, Flerov and Popov⁴⁸ published in 1933 in German the basic general principles of a soil-block test. The significance of the article by these two Russian investigators was apparently completely lost on American workers until the publication in England in 1946 of Cartwright and Findlay's "Decay of Timber and its Prevention."²⁶ Findlay had been a member of the Berlin conference. Flerov and Popov were familiar with the discussions and result of the conference, and decided in favor of the soil base for their cultures after a critical review of the various methods then in use. Their proposals to all intents and purposes were unknown here.

Van den Berge's comprehensive thesis¹¹⁶ on "Testing the Suitability of Fungicides for Wood Preservation" appeared in Dutch in 1934. A mimeographed English translation was made available soon after for limited distribution. European workers were about ready to confine the use of the agar toximetric test to determining relative toxicities only of various preservatives *in an agar medium*. Liese and his colleagues⁷¹ summarized the arguments and experiments on the agar-block method in 1935, and launched it into a status of general acceptance in Europe and Great Britain. The British^{24, 45} and German³³ editions of the standard were issued in 1939. The Rumanian version¹¹⁰ — closely following the German — came out in 1950. Jacquot,⁶⁴ Lutz⁷⁷ and Alliot² worked out proposals for standard procedures that would be more comprehensive and in their opinion better applicable to wood preservation research in France.

The Petri dish — and later the stoppered Erlenmeyer flask — agar methods continued to be used by many American investigators for testing wood preservatives, and there is no denying a certain utility in these methods for developing information about fungus poisons. The persistency of the agar techniques can be traced through publications by Richards,⁹⁴ Schmitz,^{99, 101} Snell and Shipley,¹⁰⁸ Schmitz, Buckman and von Schrenk,¹⁰² Schmitz, von Schrenk and Kammerer,¹⁰⁸ Bland¹⁴ and Hatfield.⁵³ Baechler still uses the closed flask-agar method and the fungus called Madison 517 for determining basic toximetric values;^{4, 5} and Finholt⁴⁶ has recently been bold enough to state that "Fungitoxic materials

can be evaluated as wood preservatives by mixing the toxic substances with a malt extract agar solution and then testing the mix against standard fungi."

Flerov and Popov had used sand in their preliminary experiments, but they were by no means the first to do so (see Falck⁴⁴). Rabanus⁸⁶ had reported his experiments with sand-block cultures two years earlier. He placed a pair of wood blocks — one treated and one untreated — on glass rods on wet sand in Erlenmeyer flasks; and after sterilization he inoculated the blocks directly with his test fungi. He points out that in this procedure the conditions were less favorable for the fungi than when the treated wood is placed above or on a vigorously growing culture, as in the agar-block test.

Since the papers by Rabanus and by Flerov and Popov appeared in the same journal, one can assume that the latter knew of Rabanus' work. How much any of them knew of still earlier work by Breazzano is uncertain. His work in Italy,^{20, 21, 22, 23} begun in the first decade of the century, is evidence of the intense interest of the management of the Italian railroads in some practical laboratory means for testing wood preservatives that would provide results sooner and with more definiteness than the traditional service tests. Parts of Breazzano's report of Oct. 9, 1913 are worth quoting in full from the English translation as historical background information. He reviews the situation as he sees it, and says:

"New systems and various substances for injection into wood are constantly being put on the market by industrial concerns, so that the Railway Administration finds itself confronted by an ever increasing number of processes to be examined and tested for efficiency."

By 1910 the Railway Experimental Institute

"... is well on the way toward testing the efficacy of a system of wood preservation by a method which gives dependable results even after a few months of observation.

"... after making use also of the advice on the subject received directly from Prof. Tubeuf and from Netzsche's laboratory ... positive results were obtained with the following technique:

"On the bottom of an Erlenmeyer flask of 200 ml capacity was placed a thin layer of sand." After sterilization in dry heat at 180°C "sterilized water was poured on the sand to moisten it well. Then there was placed on the sand the sample of wood, of dimensions about 9 x 2 x 1 cm., with one end resting on the damp sand and the other on the inside wall of the flask."

The whole setup was sterilized in an autoclave at 120°C for about 20

minutes. The wood was inoculated by placing a piece of a culture of *Coniophora cerebella*, grown on agar medium, directly on the wood. Breazzano states that the wood was kept moist enough because of the water in the sand, that the fungus grew luxuriantly, and that "the development of the fungus was evidently at the expense of the wood, since no other nutritive substance was at its disposal."

He used blocks cut from treated beech ties. The fungus grew readily and he concludes that the treatment was not effective. He ends this early report with the statement:

"... If the experiment is carried on under carefully defined conditions the various methods proposed for immunizing woods can be judged all by the same standard."

Breazzano presented his method at Pisa in 1919, and in 1922²¹ the principles of the sand-block culture were proposed as standard procedure (for Italy) for evaluating wood preservatives. Precise directions were given for the whole test technique, with important modification of the cultures, as indicated in the steps outlined below:

1. Sterilize by dry heat, at 180°C, "soyka" boxes 8 cm in diameter and 4 cm in height "in which is first placed a layer of sand 1 cm deep".
2. Prepare blocks of wood — treated and untreated — 4 x 4 x 2 cm, cutting them so that the broader faces will be transverse sections; and place these test blocks broad face down on the sand.
3. Sterilize at 100°C for one hour.
4. After sterilizing and cooling add sterile water in an amount that will be slightly in excess of what the sand can absorb.
5. After the wood blocks become moist plant *Coniophora cerebella* — without carrying over any agar medium with the transplant.
6. Incubate the "soyka" box cultures in a covered crystallizing dish in a dark place for one month at 20–25°C; and "Take care that in this time the water which the sand absorbs does not evaporate completely, and add sterile water when necessary."

At the end of the test the wood blocks were to be examined for decay; and if there was any doubt the wood was to be sectioned and examined microscopically for the presence of wood-destroying fungus hyphae (threads).

In retrospect the subsequent changes involving the use of soil instead of sand, and in the testing of blocks specially treated for the experiment, seem like refinements of Breazzano's methods. He later shifted to the use of very thin pieces of treated wood for his test specimens,^{22 23} severely criticizing the agar-block method that grew out of the Berlin conference as time consuming and inaccurate (loc. cit.).⁷²

In Bell Telephone Laboratories, R. E. Waterman and his colleagues started work on a wood-over-water block method for testing wood preservatives soon after the St. Louis conference, and they published their early results in 1937 and 1938.^{67, 126, 127} Their block was a $\frac{3}{4}$ -inch cube with a hole drilled through it in the approximate center of a transverse face. The $\frac{3}{4}$ -inch cubes simply represented sections of the $\frac{3}{4}$ -inch square stakes that had been substituted for round saplings⁶⁹ in the small specimen test plot experiments. The hole served a double purpose — it facilitated handling the blocks during drying and sorting operations¹²⁶ and it served as a point of entrance of moisture, which was purposely provided for the block by means of a wood wick.

Leutritz⁷⁰ formalized a soil-block test completely independently of Flerov and Popov, and published his method in this JOURNAL (Vol. 25) in 1946, following an earlier short article in 1939⁶⁸ suggesting soil as a culture medium.

Beginning in the summer of 1944 and continuing until June 30, 1951, Bell Telephone Laboratories subsidized in part a series of studies by the Madison Branch of the Division of Forest Pathology, of the United States Department of Agriculture, Bureau of Plant Industry, in cooperation with the Forest Products Laboratory at Madison, Wisconsin. The results of these studies and of parallel investigations have appeared in eight papers^{16, 35, 36, 37, 38, 39, 40, 95} from 1947 to date. The differences between the agar-block and the soil-block techniques, and the results obtained in comparable test series by the two methods are of fundamental importance. They are presented and discussed at length in a paper by Duncan.⁴¹ Already some 40,000 blocks have been tested by the soil-block method at Madison, with 75 oil-type preservatives. Both at Madison and at Bell Telephone Laboratories, Murray Hill, additional work aimed at further refining of the soil-block technique is under way.

Subsequent to discussions of the new soil-block techniques between representatives of Bell Telephone Laboratories and of the Forest Products Laboratories of Canada, Sedziak¹⁰⁶ has developed a soil-block test involving burying the block in the soil all but one corner; and instead of placing it on a fungus culture growing on feeder blocks, he inoculates a corner of the test block directly.

For a general review of laboratory and test plot methods for evaluating wood preservatives the interested reader should have available, in addition to Cartwright and Findlay's book,²⁶ at least two more recent books, namely "Wood Preservation During the Last 50 Years" by van Groenou, Rischen and van den Berge,¹¹⁸ and the third edition of *Holzkonserverung* by Mahlke-Troschel-Liese.⁷⁸ Hunt and Garratt⁶² survey wood preservation

with particular reference to the American scene. The works of Boyce,¹⁹ Baxter¹⁰ and Hubert⁵⁸ should be consulted for general information on wood-destroying fungi and the pathology of timber products. Kaufert⁶⁵ prepared a concise bibliography of pertinent articles in 1949. For a fuller coverage the book by van Groenou, Rischen and van den Berge will be found most stimulating.

Much of the European work on the testing and application of wood preservatives has been summarized in challenging form by the investigators at the Berlin-Dahlem testing station.⁵⁴ In this memorial volume, the first paper, by Schulze, Theden and Starfinger, is a compilation of the results of comparative laboratory tests of wood preservatives by the agar-block method. So much work has been done that the ingenious graphical summary table is about 12 feet long; and even then the authors have omitted many results because the conditions of the standard test³³ were not observed. Becker⁵⁴⁽²⁾ brings up to date the results of testing insecticides in the second article; Becker,⁵⁴⁽³⁾ in the next paper, summarizes tests for termite control; and Becker and Schulze⁵⁴⁽⁴⁾ in the fourth article cover laboratory tests of preservative materials for the control of marine borers. Six additional articles on subjects directly related to wood preservation complete an excellent supplement to the Mahlke-Troschel-Liese book already cited. The emphasis is, somewhat naturally, centered on the work of the Berlin station.

Rennerfelt and his colleagues^{42, 43, 88} are conducting a series of laboratory, decay chamber and test plot experiments in Sweden, aimed at evaluating wood preservatives for use in that country, and at possible correlation of experimental results with actual experience.

Bienfait and Hof¹³ are working in Holland on what appear to be the broadest test post experiments in Europe at the present time, under both land and water exposure conditions. Their tests of 10 preservatives and some 3350 posts of Douglas fir, Scotch pine, European larch, Sitka spruce, poplar and willow rival Bell Telephone Laboratories' installations in four test plots at Gulfport, Miss., Orange Park, Fla., Chester, N. J., and Limon, Colo.^{69, 75, 76} and the Forest Products Laboratory installations in Mississippi.¹⁷ Bienfait and Hof, like Rennerfelt, have been using the standard European agar-block test in their plan for correlation of laboratory and field results. No report on the Holland tests has appeared since 1948.

Narayanamurti and his associates⁸² in their first interim report on laboratory and field tests of creosotes of Indian origin present the results of some fifteen years work at the Forest Research Institute at Dehra Dun, indicating from still another quarter the compelling force that is

leading to the development of preservative evaluation methods to supplement or partly displace long and uncertain service tests. The authors present a mass of information on six different creosotes, on four creosote fractions, and on mixtures of the creosote with fuel oil of Persian origin. Sal (*Shorea robusta*) railway ties were used for the field trials. Many of the data are condensed into graphs that are small and difficult to read. The findings in general are favorable to the creosote-petroleum blends. The writer, on the basis of personal experience, is dubious about either the theoretical or practical significance, in experiments of the type reported, of the values given for standard deviations and standard errors. The scope of the work entitles it to more complete review than is practicable at this particular time and place.

Bell Telephone Laboratories are represented in a group carrying on comprehensive cooperative investigations of pedigreed creosotes on which four papers have already been published.^{5, 12, 39, 93}

The results of outdoor tests of small stakes and fence posts are issued periodically by the Forest Products Laboratory at Madison, Wis.^{15, 17, 18, 63} In this connection, the Proceedings of the American Wood-Preservers' Association are in the class of required reading. Additional references will be cited at appropriate points in the succeeding paragraphs.

Data will first be presented on some of the experience of Bell Telephone Laboratories and others with laboratory soil-block tests, with outdoor tests of small stakes, of pole-diameter posts, and with pole test lines in evaluating wood preservatives. Through analysis and discussion an attempt will be made to interpret the significance of the results obtained by the various evaluation procedures and to correlate the evidence. Emphasis will be placed naturally on creosote and pentachlorophenol because of their great importance to the Bell System pole plant. The writer intends to support his interpretations with experimental data wherever possible, reserving the privilege in some cases to make suggestions as to possible significance, even though complete technical proof may be lacking at present.

EVALUATION BY SOIL-BLOCK TESTS

General Procedures

Soil-block cultures have been described in a number of papers^{95, 96, 52, 38, 39} since Leutritz presented his method in this JOURNAL in 1946.⁷⁰ Some of the following statements, therefore, will be repetition; but the intent is to outline the technique employed at Bell Telephone Laboratories as a base for later discussion.

The culture jars are wide mouth cylindrical 8-ounce bottles, provided with screw caps. The moisture content of the soil is predetermined on a representative sample, and enough distilled water is placed in the bottle so that when the soil is added its moisture content will be somewhat above 40 per cent by weight. The bottles are filled approximately half-full of screened field top soil — which means about 140 grams of an oven-dried sandy loam. The soil handles better if it is reasonably dry so that it can be poured through a suitable funnel; and putting the water in bottles before one puts in the soil results in a practically clean glass surface on the inside of the bottle above the soil level.

Two southern pine sapwood feeder blocks, measuring $1\frac{3}{8}$ inches in the direction of the grain by $\frac{3}{4}$ inch by approximately $\frac{5}{32}$ inch ($35 \times 20 \times 4.0$ – 4.5 mm) are placed carefully on the flattened soil surface, as shown in Fig. 1(a). The soil and feeder block setups are then sterilized for one-half hour at a pressure of 15 pounds per square inch, after which they are allowed to cool in the autoclave.

Inoculation is accomplished by carefully placing a piece of inoculum, cut from a fresh Petri dish culture of what may be called a standard test organism, at or near the middle of the feeder block surfaces. Under

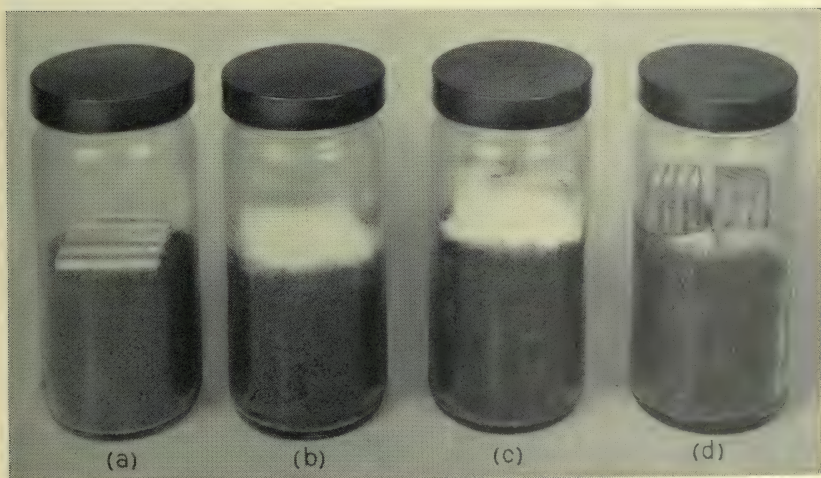


Fig 1—Four eight-ounce cylindrical bottles illustrating the soil-block cultures: (a) Bottle half-full of top soil, containing 40 per cent moisture on an oven-dry soil basis, with two flat feeder blocks of southern pine sapwood on top. (b) A sixteen-day old culture ready to receive the impregnated southern pine sapwood test blocks. (c) Two laboratory weathered test blocks from the same series treated to a below threshold concentration of pentachlorophenol 0.051 lb dry penta/cu ft, attacked by the test fungus, *Lenzites trabea*. (d) Two test blocks, laboratory weathered, treated to a retention of 0.194 lb dry penta/cu ft, near the threshold retention, showing resistance to fungus attack.

the temperature and humidity conditions of the incubation room, the growth of the fungus mycelium covers the feeder blocks in about two weeks and the fungus threads are then well started downward into the soil.

Treated test blocks are weathered and then conditioned under controlled temperature and humidity to approximate constant weight. They are then sterilized, along with untreated control blocks, in an autoclave for 15 minutes at 100°C, atmospheric pressure.

As a rule two treated blocks having approximately the same retention of preservative are placed together in a single test bottle. The incubation period is three months, in an incubation room held at a temperature of $80 \pm 2^\circ\text{F}$ and at a relative humidity of 70 ± 2 per cent. At the end of this period the cultures are taken down. This means that the blocks are removed from the bottles, brushed free of fungus mycelium, and weighed immediately. They are given a preliminary examination for decay evidence, and then reconditioned, under the same temperature and humidity conditions as before sterilization, to approximate constant weight. Fungus attack is determined by observation and by weight losses. The general setup of the cultures is illustrated in Fig. 1 (a-d).

Inoculation and Incubation Rooms

To facilitate handling the soil-block cultures, an inoculation room and an incubation room have been built (Fig. 2) at the Murray Hill Laboratories. Both are held at approximately the same temperature and relative humidity, that is, 80°F and 70 per cent. The inoculation room serves as a lock chamber, and passage from it to the incubation room has a negligible effect on the humidity and temperature of the incubation room. The latter is provided with an illuminated double plate glass window (Fig. 3), so that the interior can be exhibited without the necessity of entering the room. This window is fitted with a heavy roller shade, and the room ordinarily is kept dark.

Soil Characteristics and Moisture Content

The question that is asked most often about the cultures is whether a standard soil is used. European and American criticism has been definitely directed^{78, 122} at the fact that the use of different soils might have so much effect upon the growth and the reaction of the test fungi in the cultures that quite different results would be obtained by investigators in different laboratories. This possibility is recognized; but the evidence to date seems to point to the general conclusion that perhaps the prin-



Fig. 2—Soil-block cultures on the shelves in the incubation room. The unpainted wood shelves come to equilibrium with the temperature and relative humidity and thus are a factor in keeping the conditions stable. The back edges of the shelves are set away from the wall to provide spaces for air circulation. The front edges of the shelves are provided with metal labeling strips.



cipal and most important factor in the soil-block culture is the moisture holding capacity and content of the soil, rather than its nutrient function. If continued experimentation supports this conclusion it would not be necessary to limit the type of soils used except within rather broad limits. It also appears that the size and thickness of the feeder block now employed introduces enough wood into the culture bottle to mask any minor variations in the soil itself. The all important thing is to have enough water in the soil *throughout the test period* to keep the air above the soil essentially at 100 per cent humidity and the blocks at about fiber saturation — say about 27 per cent, oven-dry weight basis.

The soil in use at Bell Laboratories at present is obtained from a plot that has been set aside at the Chester (N. J.) Test Station. This plot has been fallow for twenty-five years. It supports a general grassy flora. The soil is a sandy loam with the following general description:

pH.....	4.9-5.0
Available magnesium.....	37.5 lb/acre
Available phosphorus.....	4.5 lb/acre
Available potassium.....	70.0 lb/acre
Organic matter.....	3.0 per cent

The cultures at the Forest Products Laboratory⁴¹ have been made with a silt loam having a pH between 5.5 and 6.0. Bell Telephone Laboratories' tests have indicated the desirability of avoiding soils of either very sandy or very heavy clay types. The soil from the Chester Test Station described above is being used in all cultures, and there is a sufficient layer of top soil on the reserved plot to make parallel cultures for a good many years. Until such time as more definite and positive information on the effect of minor variations in the soil type are determined it is generally agreed that all of the comparative tests in any given series at least should be run on the same soil. Experimental work is now under way to determine the possible advantage of the addition of Krilium* to the soil in the culture bottles to maintain porosity and an even, high moisture holding capacity.

After the test blocks are placed in the culture bottles and during the course of the ninety-day incubation period the screw caps are left *loose*. The general technique followed in making up the soil cultures, as far as moisture is concerned, parallels that used at Madison. The moisture content is close to that recommended by Flerov and Popov,⁴⁸ namely 40-50 per cent of the weight of the soil plus the feeder block, with distilled water added during the test period, if necessary to maintain good

* An acrylonitrile product of the Monsanto Chemical Company.

growth of the test fungus. Breazzano²¹ thoroughly saturated the sand base in his test cultures. Leutritz⁷⁰ and Harrow^{51, 52} working with tightly closed culture jars found a 25 per cent level in the soil to be satisfactory.

Flerov and Popov state after special control tests "that replacement of (the) sand by soil had no effect on the results of the tests and only shortened their duration." Duncan⁴¹ has found from her tests that variations in moisture content and soil type affect the degree of fungus attack only and that they do not change the determination of the treatment threshold concentration in any given set of test blocks.

Even-Aged Cultures

The thickness of the feeder blocks has been gradually increased to about $\frac{3}{16}$ inch, or between 4 and 4.5 millimeters. This provides food for the fungus to establish itself in the bottle. The inoculum pieces are roughly 1 cm square, cut from Petri dish cultures that are 15 ± 1 days old. The planting routine is carefully scheduled so that even-aged soil-block cultures — 13–15 days — are ready to receive the treated blocks when the latter are ready to be placed in test. This principle of using even-aged cultures has been stressed by the Madison investigators, and it is considered to be a factor of major importance in the proper culture technique.

Standard Test Organisms

There have been continuous discussions since the beginning of laboratory tests in Europe, as well as in this country, about what test organisms should be used. Conforming to the experience and practice at Madison the following three numbered strains of wood-destroying fungi are recognized as the "standard" strains for the testing of oil type preservatives in coniferous wood:

Lentinus lepideus, Madison 534
Lenzites trabea, Madison 617
Poria monticola, Madison 698

All three are known to be associated with the decay of treated timber. *Lentinus lepideus* is particularly tolerant of creosote,^{41, 66} and relatively susceptible to pentachlorophenol. It has frequently been isolated from decaying creosoted southern pine poles and other creosoted coniferous timber in contact with the ground. *Lenzites trabea* is generally an "above ground" fungus. It also has been isolated from decaying creosoted timber; and it is the principal cause of "shell rot" in the above ground sap-

wood of western red cedar poles. It is relatively susceptible to creosote and quite tolerant of pentachlorophenol in the block tests. *Poria monticola* is relatively tolerant of pentachlorophenol and of copper compounds under laboratory test conditions, and relatively susceptible to creosote. It is of special interest also because it may be identical with some of the fungi tested in Europe under the name of *Poria vaporaria*, and thus its use may facilitate comparisons of a sort with results obtained by other investigators. For instance, information has reached the Division of Forest Pathology at Madison, from Findlay at the Princes Risborough laboratory in England, that Harrow's *Poria vaporaria*⁵¹ is the same as Liese's,⁷¹ and that it has been identified as a strain of *Poria monticola* by Miss M. Nobles of Canada.

Within the last few years another fungus, characterized by the formation of conspicuous saffron yellow strands, has been found associated with decayed specimens of creosoted pine poles.⁶⁰ The writer has seen the tell-tale strands in old cull dumps only. It has been identified as *Poria radiculosa*. Whether it is truly a primary attacker or a secondary organism is not yet clear. Soil-block tests are under way at Madison to determine its significance as a possible species to supplement *Lentinus lepideus* in the evaluation of creosote.

In connection with the use of the three numbered "standard" strains listed above, there may always be some reasonable doubt as to whether the cultures employed in different laboratories have the same virulence. To answer this question precisely involves a lot of careful biological check testing, and such tests are already being made in the Division of Forest Pathology at the Plant Industry Station, Beltsville, Md. It is assumed for the time being that the numbered strains are virulent and satisfactory test organisms for such preservatives as creosotes and pentachlorophenol-petroleum solutions.

The Scope of the Soil-Block Evaluation Test

For a complete understanding of the scope of the soil-block evaluation test it is necessary to consider this test as having two functions. The first function involves the use of the soil-block test per se (without weathering) to measure the reaction of the test organisms to various quantities of a given preservative, and to compare these reactions against different preservatives. In this function the test has been used in lieu of the agar Petri dish test,⁹⁴ and is considered to be much more satisfactory as a screening test by workers at the Laboratories. It has been employed at Madison for testing the natural durability of wood, plywood, fiber board, etc.

The second — and more important — function of the soil-block evaluation is that incorporating a weathering or aging procedure. This puts the test in the more practical category of testing the wood preservative properties, viz., *toxicity and permanence*. In this respect it has something in common with the German Standard DIN DVM 2176³³ for short time mycological testing of wood preservatives by the block method and covers a broad concept from the treating through partial aging of the blocks. Separate German standards cover procedures for leaching³⁴ and volatility tests,⁷⁸ (p. 264 and Fig. 42 of Reference 78).

The Laboratories' concept of the scope of the soil-block test including the weathering procedure is definite. It must be appreciated that this method which employs manipulative procedures involving both toxicity and permanency yields significant data in a period of only a few months. The data derived must be correlated subsequently, of course, with the data covering the results of tests of $\frac{3}{4}$ inch stakes six to seven years later, with the data on test posts some ten to fifteen years later, and with data on poles in line some twenty-five years later.

It is only being realistic to say that the Bell System cannot afford to wait for physical life tests of new materials under natural conditions of exposure before recommending them where techniques and extensive experience permit acceptable estimates to be made from accelerated evaluation in relatively short periods of time.

Preparation of the Test Blocks — Manufacture

Southern pine sapwood, free from stain or decay, is used as a base material for the test blocks. The process of manufacture begins at the saw mill, where freshly cut logs selected for the purpose are carefully sawed into one inch boards. Straight grain material is most desirable. The boards are kiln-dried immediately and shipped as soon as practicable to the Laboratories. It has been the practice to store the boards in a steam heated basement where the humidity is low enough to hold the moisture content of the boards down to about 5 to 7 per cent. The sapwood only is used, which means that any small heartwood portions must be marked out for rejection. The blocks are accurately cut $\frac{3}{4}$ -inch cubes. A $\frac{1}{8}$ -inch hole is drilled through the center of the tangential surfaces of each block. It has been found that drilling the hole through the transverse surface, which was the early practice with Waterman, Leutritz and Hill¹²⁶ is a difficult procedure; and sometimes it amounts to an impossibility because the harder summerwood layers deflect and break the drills. In any event, drilling through the tangential surface opens up more paths for longitudinal absorption and penetration, as well as

evaporation, of the preservative. The feeder blocks and the $\frac{3}{4}$ -inch test blocks are usually made at the same time, from parts of the same boards. The blocks are kept clean, and reserve stocks are carefully stored in a dry room. The blocks in storage reach an approximate moisture equilibrium of 6 to 7 per cent, on an oven-dry weight basis.

Test Block Selection for Density

Random samples of the blocks are weighed and segregated into groups at 0.1-gram intervals, 4.10 to 4.19 grams and 4.20 to 4.29 grams, for example. Blocks of practically equal weight can be chosen for the comparison within any given series of different concentrations of a preservative. The weighed groups of blocks are kept in convenient lot sizes in a dry place. Since the blocks are accurately cut the segregation by weight amounts to a segregation by density.

It has not been found necessary or practicable to separate the blocks into groups with the same numbers of annual rings, although in some instances an approximation to this ideal has been attempted. Furthermore, it has not been found practicable to separate the blocks on the basis of the direction in which the rings run across their transverse faces. From experience to date it does not appear that either ring direction or ring count has any material effect upon the behavior of the blocks in the culture as far as determination of preservative thresholds are concerned; but experiments are under way at Madison to determine the effect of density on the relative degree of decay. Inasmuch as all of the blocks are placed in culture with the transverse surface down, so that alternate spring- and summerwood layers are exposed directly to the test organism, the latter can enter either springwood or summerwood in accordance with its ability to resist the concentration of the preservative present in these two parts of the annual ring.

Average Block Volume

The average volume of the oven-dried blocks, determined from random samples by a mercury displacement technique, was found to be 6.484 cc, with a standard deviation of 0.0831. This represents a coefficient of variability of 1.28 per cent. The minimum-maximum range of volumes ran from 5.93 cc to 6.87 cc. These extreme deviations are normally detected in handling the blocks and both high and low volume blocks are rejected. The variation in density and volume of the test blocks will be discussed separately in the paragraphs dealing with the treatment of the blocks.

Treatment of the Test Blocks

The blocks selected for any given treatment are numbered serially with India ink on the upper half of one of the radial faces. All blocks are then oven-dried for 24 hours at 105°C to an approximate constant weight. The blocks are removed from the oven, and placed in a desiccator over P_2O_5 . Check tests of blocks held under these conditions show that they do not change weight by more than one hundredth gram within the period they are held for weighing. The cooled, oven-dried blocks are weighed to the nearest hundredth gram.

The weighed blocks are placed in beakers and arranged with a tangential face down so that the transverse surfaces do not touch and the holes are vertical. This refinement in placing the blocks may not be necessary to obtain satisfactory absorption, but the procedure has worked out well, and it has been followed consistently. For any given concentration a sufficient amount of creosote, for example, and toluene are combined by weight to leave in the blocks, after treatment, the desired retention of preservative. Experience has indicated the concentration required, which depends to a certain extent upon the type of vacuum equipment that is available as well as upon the density of the blocks to be treated and the nature of the treating solution. Actually the process of treatment is simple. The beaker containing a given lot of weighed oven-dried blocks is placed within a bell jar and subjected to a vacuum of 3 to 4 millimeters of mercury. When this vacuum has been reached the line to the vacuum pump is shut off, and the preservative is run into the beaker from a separatory funnel¹²⁶ fitted into a rubber stopper on the top of the vacuum chamber, the blocks being weighted down below the level of the preservative.

The absorption and distribution of the oil within the blocks seems to take place very rapidly. Generally speaking, the beaker containing the blocks and the preservative are removed from the vacuum chamber as soon as practicable to permit continuing the treatment of another group of blocks. However, the blocks are usually held in the preservative solution for an hour or two, which apparently is long enough to bring about essentially complete saturation. When all the treatments for a given group of concentrations have been finished (a) the treated blocks are wiped to remove the excess oil, and (b) they are weighed immediately to 0.01 gram. The retentions of creosote or pentachlorophenol, for example, are determined on a gain in weight basis by calculations from the amount of material picked up during the treatment and the concentration of the preservative in the treating solution.

Retention Gradients

In the hope of setting at rest some of the doubts and criticisms that have arisen about the accuracy of the treatments and the retention of preservative in the blocks, some results of the treatment process just described will be presented in rather elaborate detail.

The success of the treatments depends upon experience, as indicated previously, with the particular type of vacuum equipment available. However, once the level of performance to be expected from the vacuum equipment is learned, one has to take into account the variations that are introduced by the density of the blocks and by the specific gravity of the treating solution. It is the intent in all of the treatments at the Laboratories to arrive at a series of gradient retentions, on as accurate a line as possible, and as nearly as possible equal gradients, so that the fairest comparison can be made of the behavior of the different preservatives. Fig. 4 shows the gradient obtained by plotting the data shown in Table I for retention of creosote and retention of pentachlorophenol solution over the concentration of these preservatives in the treating solution. The analysis of the creosote — BTL 5340 — is shown in Table II. The slopes of the two gradients are considered to be about as close as the experimental procedure will permit. Fig. 5 shows the gradient

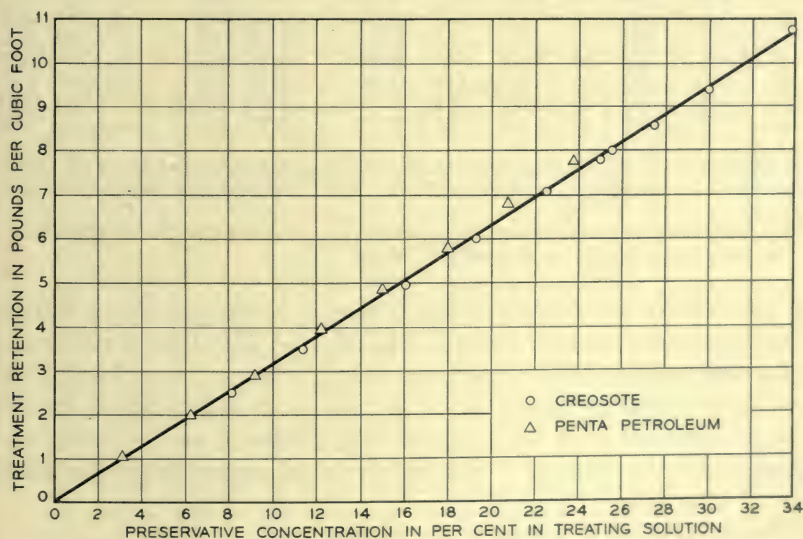


Fig. 4—Gradient retentions for comparative soil-block tests of a creosote (BTL No. 5340) and a penta-petroleum solution (4.92 per cent pentachlorophenol in Standard Oil Company of New Jersey No. 2105 Process Oil). The preservatives were used in toluene solution.

TABLE I — FULL-CELL TREATMENT

Soil block tests with creosote (No. 5340, see Table II) and with pentachlorophenol-petroleum (4.92 per cent in Standard of New Jersey No. 2105 Process oil) in toluene; absorption and retention of preservative data for parallel comparative tests.

Charge No.	Average Oven-dry		n	Average Absorption*		C†	Average retention							
	Weight	Density		Total	Per cc		Whole preservative			Creosote		Penta		
							Creo- sote	Penta						
								Sol.	Penta					
										Total	Per cc	Total	Per cc	
	(gms)		(gms)	(per cent)	(lb/cu ft)	(gms)	(gms)							
9	3.77	.584	30	3.28	.509	8.18	2.49	—	—	.28	.043	—	—	
10	3.78	.586	30	3.40	.527	11.45	3.48	—	—	.39	.061	—	—	
6	3.80	.589	30	3.44	.533	16.11	4.96	—	—	.55	.085	—	—	
5	3.83	.594	30	3.46	.536	19.33	5.99	—	—	.67	.104	—	—	
8	3.80	.589	30	3.51	.544	22.55	7.08	—	—	.79	.123	—	—	
1	3.74	.580	30	3.49	.541	25.00	7.81	—	—	.87	.135	—	—	
4	3.77	.584	30	3.51	.544	25.55	8.03	—	—	.90	.140	—	—	
3	3.80	.589	30	3.49	.541	27.50	8.58	—	—	.96	.149	—	—	
2	3.78	.586	30	3.50	.543	30.00	9.40	—	—	1.05	.163	—	—	
11	3.70	.574	30	3.33	.516	3.25		1.05	.052	—	—	.005	.0008	
12	3.79	.588	30	3.33	.516	6.25		2.01	.099	—	—	.010	.0016	
13	3.79	.588	30	3.29	.510	9.25		2.95	.145	—	—	.015	.0023	
14	3.75	.581	30	3.31	.513	12.25		3.96	.193	—	—	.020	.0031	
15	3.71	.575	30	3.34	.518	15.00		4.84	.238	—	—	.025	.0039	
16	3.70	.574	30	3.34	.518	18.00		5.81	.286	—	—	.030	.0047	
17	3.72	.577	30	3.38	.524	23.75		6.79	.334	—	—	.035	.0054	
18	3.73	.578	30	3.38	.524	23.75		7.77	.382	—	—	.040	.0062	

* Absorption is the total amount of the treating solution picked up at treatment, that is, the gain in weight, including both preservative and the toluene carrier.

† C is the concentration of the preservative, e.g., creosote or penta petroleum, in the treating solution, in grams per 100 ml.

for pentachlorophenol alone, without regard to the petroleum carrier, also plotted from data in Table I. The scale on the abscissa represents the concentration of either the creosote or the pentachlorophenol solution. The ordinate represents pounds per cubic foot retained by the blocks, calculated from the pickup during treatment and from the concentration of the creosote or pentachlorophenol in the treating solution.

The Amount of Preservative in the Blocks

The use of these gradient concentrations is a continuation of the procedure worked out in the earlier stages^{39, 41} of the Madison tests.

TABLE II — ANALYSES OF CREOSOTE BTL No. 5340,
WATER-FREE BASIS

	1. Fall, 1946	2. Spring, 1952*
Specific gravity 38/15.5°C.....	1.088	1.102
Distillation, per cent, cumulative		
to 210°C.....	0.00	0.00
210-235.....	0.80	0.00
235-270.....	12.87	13.59
270-300.....	42.12	
300-315.....	54.30	52.03
315-355.....	79.10	78.05
Residue above 355°C.....	20.90	21.64
Total.....	100.00	99.69
Sulph. res., gm/100 ml.....	0.51	0.59
Tar acids, gm/100 ml.....	4.10	4.44
Benzol insol., per cent.....	0.07	0.59
Specific gravity (38°C) 235-315°C.....		1.053
315-355°C.....		1.118

* Average of 2 analyses.

It should be noted that the retentions are calculated as averages for the respective charges. Attention is called to the small quantity of preservative material involved. Even in calculated retentions of creosote, for example, 9.40 pounds per cubic foot (Table I), the retention means 1.05 grams in the whole block, or 0.163 grams *in each cc of block volume*. In

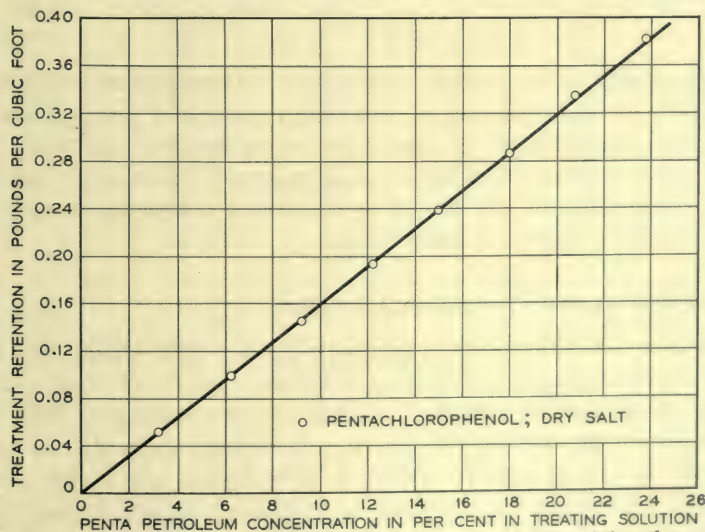


Fig. 5—Gradient retention of pentachlorophenol, calculated for the material alone, without the oil carrier. See Fig. 4.

the highest retention employed for the penta petroleum solution the calculated net retention averaged 7.77 pounds of the penta solution per cubic foot, or 0.382 pounds of pentachlorophenol per cubic foot; and these figures represent, respectively, 40 milligrams of pentachlorophenol in the average block, or 6.2 milligrams *per cc of block volume*. Exact data on treatment are discussed in the following paragraphs. The use of carefully calculated gradient retentions in each case makes it possible to detect any wide variation in the normal behavior of the blocks either with respect to pickup during treatment or in the reaction of the test fungus to the preservative.

Data are included in Table I on average oven dry weight of the blocks,

TABLE III — FULL-CELL TREATMENTS

Soil block tests; treating solution components, per cent by weight. (See Table I).

Charge No.	Penta-petroleum	Creosote	Toluene
19	4.65	4.65	90.70
20	10.50	10.50	79.00
21	15.00	15.00	70.00
71	—	23.00	77.00
72	—	23.00	77.00
73	—	24.75	75.25
74	—	25.50	74.50
75	—	30.00	70.00
76	—	32.00	68.00

average density on an oven dry weight and volume basis, average pickup of creosote or penta solution in pounds per cubic foot and in grams per block, the concentration of the preservative materials in the toluene preservative solution, and the average grams of preservative per cc of block volume. All of the blocks in these two groups of charges were chosen within a narrow density range.

Block Density and Preservative Absorption

It will be noted that in the charges in Table I there is a general trend upward in the grams absorbed at treatment per cc of block volume. as the specific gravity of the treating solution increases. This is, of course, one of the results of increasing the concentration of creosote, for example; and furthermore, as would be expected, the higher gravity solutions represented by the creosote treatments show a higher pickup in terms of total grams as well as in grams per cc. The make-up of the

TABLE IV — FULL-CELL TREATMENT WITH PENTA-PETROLEUM-CREOSOTE IN TOLUENE

Relation of variable block density to absorption of treating solution, grams per cc of block volume; density and volume on oven-dry basis. (See Tables V and VI).

Charge 19 Av. retention 2.99 lb/cu ft		Charge 20 Av. retention 6.69 lb/cu ft		Charge 21 Av. retention 9.59 lb/cu ft	
Density oven-dry	Absorption gms/cc of block vol.	Density oven-dry	Absorption gms/cc of block vol.	Density oven-dry	Absorption gms/cc of block vol.
.440	.599	.468	.589	.484	.581
.459	.583	.483	.584	.489	.576
.459	.586	.533	.549	.505	.564
.475	.575	.543	.542	.520	.550
.507	.561	.544	.527	.529	.555
.517	.545	.557	.537	.541	.541
.533	.544	.564	.529	.549	.538
.535	.525	.567	.530	.554	.532
.543	.530	.569	.569	.555	.526
.543	.537	.604	.510	.573	.538
.571	.508	.606	.511	.582	.528
.574	.524	.611	.504	.582	.530
.608	.488	.616	.499	.592	.517
.609	.495	.618	.491	.602	.512
.611	.500	.618	.497	.610	.497
.611	.533	.624	.503	.612	.507
.623	.483	.625	.493	.614	.502
.624	.493	.627	.492	.614	.508
.637	.476	.628	.488	.615	.540
.638	.477	.644	.481	.617	.495
.641	.474	.645	.488	.617	.504
.645	.478	.646	.482	.627	.495
.646	.473	.651	.482	.633	.497
.657	.462	.662	.433	.647	.488
.660	.453	.674	.461	.650	.487
.663	.466	.676	.465	.664	.416
.663	.468	.674	.452	.672	.473
.668	.462	.690	.456	.684	.470
.691	.447	.703	.445	.687	.464
.695	.450	.738	.427	.723	.442
<i>Average</i>					
.592	.507	.614	.501	.598	.512
<i>Standard deviation</i>					
.0726	.0435	.0616	.0405	.0591	.0370
<i>Coefficient of variability—per cent</i>					
12.26	8.58	10.03	8.08	9.88	7.23

treating solutions for Charges 19–21 and 71–76, inclusive, are shown in Table III. The relation of the density of the blocks to the pickup, i.e., the absorption at time of treatment, is illustrated in Tables IV, V and VI and in Fig. 6. The data have been split up to facilitate reference.

Table IV shows the complete data for oven dry density and for absorption in grams per cc of block volume for Charges 19, 20 and 21, with values for the average, for the standard deviation, and for the coefficient of variability. The pickup varies inversely as the density, which is to be expected when random blocks instead of selected density blocks are employed. The coefficient of variability in the density figures is evidently greater than it is in the pickup figures; and a lower figure for the latter is related to a lower figure for the former.

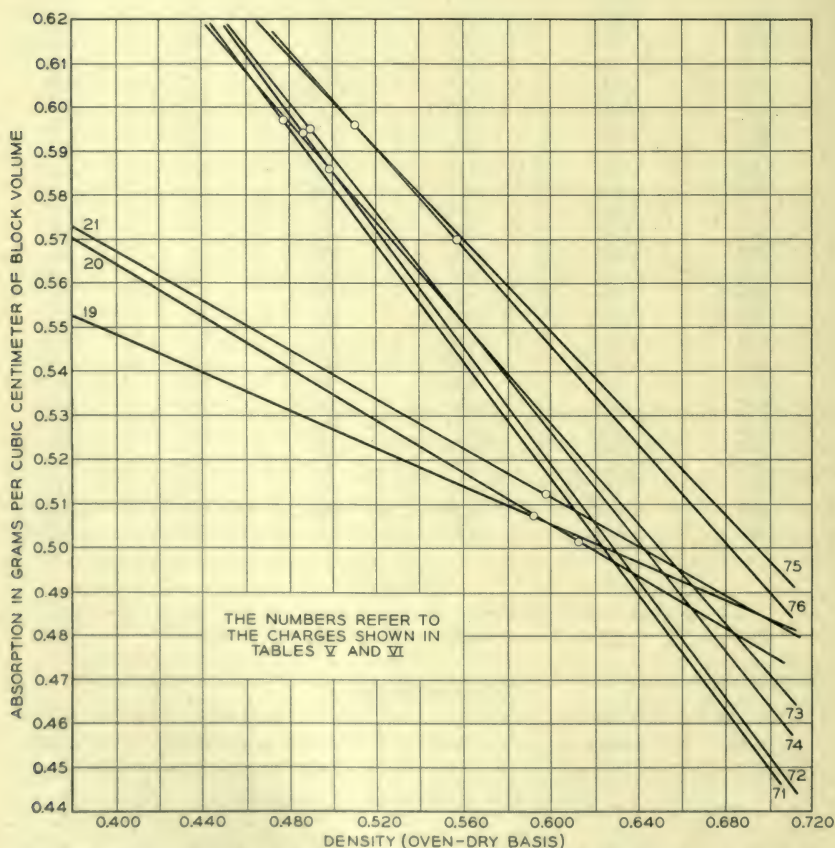


Fig. 6—Regression lines for absorption at treatment, in gms/cc of oven-dry block volume, on oven-dry density of the $\frac{3}{4}$ -inch cube test blocks.

These same values for these three charges, 19, 20 and 21, and similar values for charges 71-76, inclusive, are incorporated along with average and range of retention data in Table V, and with statistical data for regression lines in Table VI. The data serve to illustrate the degree of variability in treatment results that may occur when random blocks are used. The best indices of these variations are in the columns showing

TABLE V — RETENTION DATA FOR FULL-CELL TREATMENT

Soil block tests; average and range of preservative retention, by charges, at treatment.

Charge No.	n	Penta petroleum lb/cu ft	Pentachlorophenol						Creosote					
			lb/cu ft			gms			lb/cu ft			gms		
			Av.	Min.	Max.	Av.	Min.	Max.	Av.	Min.	Max.	Av.	Min.	Max.
19	30	1.51	.093	.064	.097	.008	.007	.010	1.48	1.29	1.97	0.154	0.134	0.205
20	30	3.35	.163	.136	.219	.017	.014	.023	3.34	2.77	4.44	0.348	0.288	0.462
21	30	4.79	.237	.192	.269	.024	.020	.028	4.80	3.90	5.46	0.500	0.407	0.569
71	30	—	—	—	—	—	—	—	8.58	7.65	9.04	0.892	0.817	0.941
72	30	—	—	—	—	—	—	—	8.52	7.67	9.20	0.886	0.798	0.957
73	30	—	—	—	—	—	—	—	9.06	8.43	9.57	0.943	0.876	1.025
74	30	—	—	—	—	—	—	—	9.46	8.61	10.27	0.984	0.895	1.068
75	30	—	—	—	—	—	—	—	11.12	10.07	12.06	1.153	1.047	1.254
76	30	—	—	—	—	—	—	—	11.37	11.05	11.94	1.184	1.149	1.242

TABLE VI — FULL-CELL TREATMENT

Soil-block tests with (a) penta-petroleum creosote, and with (b) creosote, in toluene; relation of variable block density to absorption of treating solution.

Charge No.	n	Av. retention lb/cu ft		Density oven-dry			Absorption gms/cc vol.			Correl. coeff. r	Regression of absorption Y on density X
		Creosote	Penta-petro Creosote	Av.	σ^*	c.v.†	Av.	σ^*	c.v.†		
19	30	—	2.99	.592	.0726	12.26	.507	.0435	8.58	-.3444	.6286-.2066X
20	30	—	6.69	.614	.0616	10.03	.501	.0405	8.08	-.4282	.6736-.2820X
21	30	—	9.59	.598	.0591	9.88	.512	.0370	7.23	-.4718	.6893-.2957X
71	30	8.58	—	.477	.0428	8.97	.597	.0293	4.91	-.9589	.9109-.6580X
72	30	8.52	—	.486	.0457	9.40	.594	.0326	5.49	-.9233	.9143-.6589X
73	30	9.06	—	.499	.0456	9.14	.586	.0275	4.69	-.9426	.8701-.5692X
74	30	9.46	—	.489	.0402	8.22	.595	.0279	4.69	-.8916	.8969-.6181X
75	30	11.12	—	.510	.0538	10.55	.596	.0302	5.07	-.9206	.8595-.5170X
76	30	11.37	—	.557	.0095	1.71	.570	.0114	2.00	-.4583	.8781-.5543X

* σ = standard deviation.

† c.v. = coefficient of variability, per cent.

the average and total spread in grams of preservative absorbed. The effect of selection for density is shown clearly — within the particular treatment groups — by the figures for charges 75 and 76. In the latter the use of selected even density blocks reduced the spread to below half of that in charge 75, and reduced the coefficient of variability by two-thirds.

Regression lines for pickup in grams per cc of block volume on oven dry density are shown for the two groups of charges in Fig. 6. The flatter slope of the lines for charges 19, 20 and 21 seems to reflect the difference in specific gravity and viscosity of the treating solutions. Higher densities have greater effect on absorption of higher gravity solutions. The fact remains that there is considerable uncertainty as to the significance of strict selection of the blocks for density if one uses a series of closely spaced gradient retentions.

Data for 8-pound charges for comparative soil-block tests of two of the cooperative creosotes, Nos. 7 and 9, a low residue domestic oil, low in tar acids and naphthalene, and a British vertical retort tar creosote, respectively, are condensed in Table VII. The differences in the treatment results are not considered to be significant.

Weathering

The blocks remain on the racks on the laboratory tables for about one week, which is long enough to permit the evaporation of most if not all of the toluene. Experiments have shown that when blocks are treated with toluene alone the toluene is all lost, on a weight basis, within 24 hours. When treating solutions of creosote in toluene are used the evaporation rate, also determined by weight loss, is slower; but it is believed that all of the toluene is gone before the blocks are ready for test. In any event, after the above-mentioned preliminary drying period the blocks are handled in a manner that differs somewhat from the procedure that

TABLE VII — RETENTION FOR FULL-CELL 8-POUND
TREATMENT DATA

For parallel comparative soil-block tests of cooperative creosotes No. 7 and 9 against *Lentinus lepidus*, Mad. 534: toluene-creosote treating solution.

Creosote No.	n	Density oven-dry	Creosote lb/cu ft				Creosote grams				Standard deviation	Coefficient of variability, per cent
			Avg.	Min.	Max.	Spread	Avg.	Min.	Max.	Spread		
7	30	.606	8.11	7.75	8.51	0.76	.848	.809	.889	.080	.0201	2.37
9	30	.601	7.90	7.49	8.56	1.09	.825	.780	.893	.113	.0237	2.87

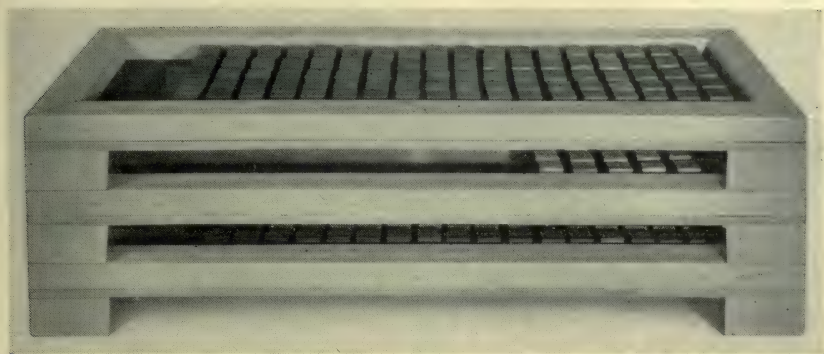


Fig. 7—Three unpainted stacked handling trays. The trays have plastic screen bottoms on which the blocks can be arranged with free air space all around, to promote even drying conditions.

has been followed up to this time at Madison. The principal difference is the omission of any tests of unweathered blocks. Instead, the emphasis is placed on the development of a weathering or aging cycle that will bring about total overall preservative losses like those that occur in $\frac{3}{4}$ -inch stake specimens or in pole-diameter posts in the Gulfport test plot. The character and extent of such losses will be discussed later.

Two systems of weathering have been employed up to the time of this writing. The first consists in soaking the blocks over the week in water that is changed morning and night and drying them at room temperature over the weekend, in accordance with German standard for leaching;³⁴ and the second is the same method that is employed in the Madison tests^{39, 41} in which the blocks are strung on nylon thread, separated by glass beads, and exposed to outdoor weathering under natural conditions for sixty days. The duration of this outdoor test has been limited to sixty days, regardless of the season or month of the year. The effectiveness of the climatic conditions at the Chester Field Station during the period from October, 1951, to April 1, 1952, compared with the roof weathering as conducted at Madison with the same creosote sample remains to be seen.

As for the German standard leaching procedure,³⁴ experience up to April 1, 1952, indicates that the method does not result in the removal of creosote, for example, in the same degree or manner as preservative materials of this type are removed by outdoor weathering conditions. It definitely is not comparable with the latter in its effect. The German leaching procedure simply uses too much water and not enough air and heat; and Bavendam⁹ quotes Falck as saying that creosote is insoluble in water and that it cannot be *washed* out of wood. The failure of the



Fig. 8—Creosoted test blocks, arranged on pins on a metal rack in a large glass tube to facilitate handling during sterilization and the subsequent operation of planting in the soil cultures.

leached creosoted blocks to decay in the reported soil-block tests on the cooperative creosotes³⁹ confirms European experience. Schulze, Theden and Starfinger^{56 (1)} (p. 15, Tab. 13 of Reference 54) indicate that there is little if any reduction in the preservative value of creosote as a result of their standard leaching tests. Therefore, in order to accelerate the weathering process new techniques are being worked out at the Laboratories in which the wet cycle is shortened (Cf. Rhodes et al,^{50, 89}) and in which, without the use of a wheel, controlled artificial heat is used to speed up evaporation during the rest of the cycle.

Conditioning

Convenient unpainted wood trays (Fig. 7) with plastic screen bottoms are used for handling the blocks in groups at any time during their processing schedule. At the end of the weathering cycle, indoor or outdoor, the blocks are arranged on such trays and conditioned to an approximate constant weight and about 12 per cent moisture content on shelves in the 80°F and 70 per cent relative humidity of the incubation room. Weights before test, to the nearest 0.01 gram, are taken at the end of the conditioning period. The relative amounts of wood, water and creosote in the blocks are determined by weight from test blocks, and by weight and extraction from control blocks after sterilization.

Sterilizing

The test blocks are arranged for sterilization on metal racks in large glass tubes (Fig. 8). The autoclave temperature is held at 100°C for 15 minutes.

Flow Chart for the Bioassay Test

The various steps in the whole evaluation procedure are indicated in the flow chart shown in Fig. 9. Rhodes⁸⁹ used a similar idea to illustrate his procedure.

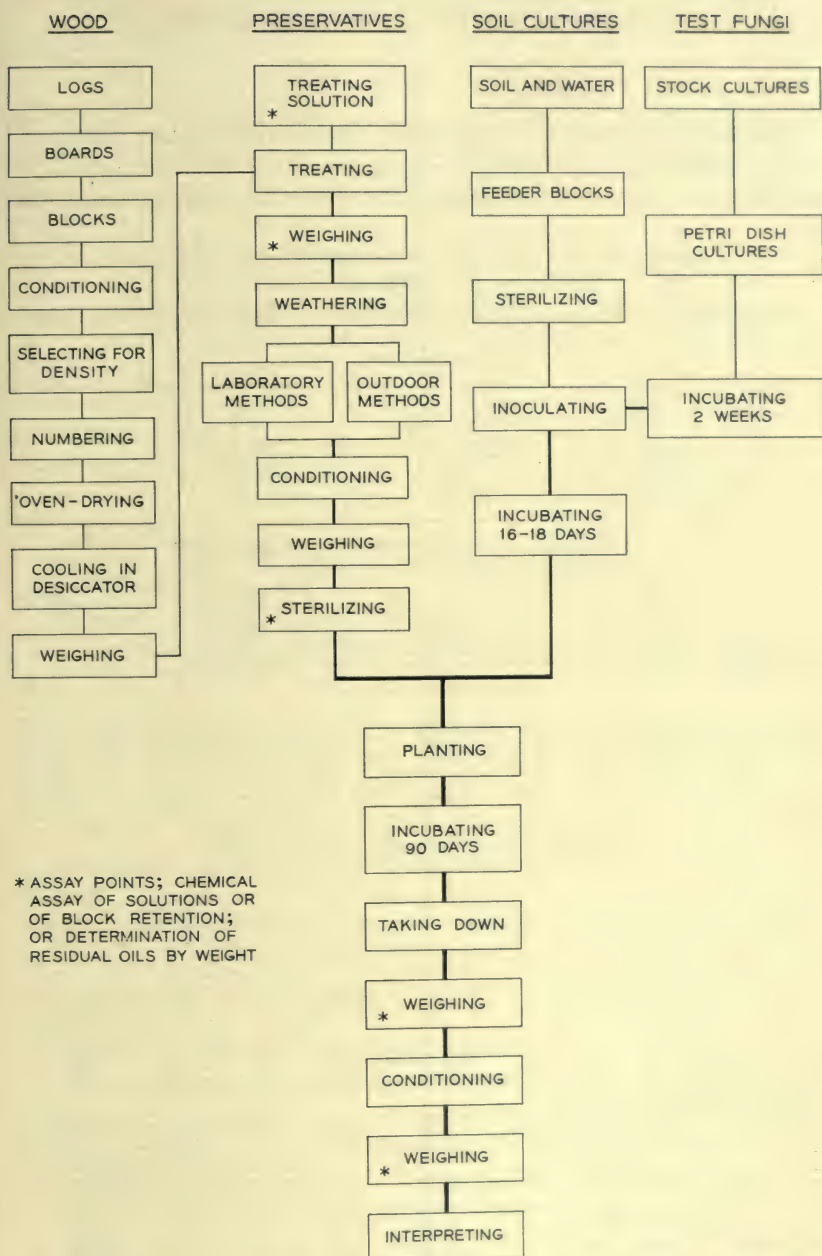


Fig. 9—Flow chart of the laboratory bioassay test procedure. Steps in the overall processing must be carefully scheduled to provide adequate time for manipulation and to assure the requisite quantity of even-aged cultures at the time the treated and weathered test blocks are ready to go into test.

Some Madison Test Results

Inasmuch as results from recent Bell Laboratories' bioassay tests will not be immediately available, data from the Madison experiments⁴¹ are used in order to illustrate the results that one may hope to secure from

TABLE VIII — BIOASSAY BY SOIL-BLOCK TESTS ON
OUTDOOR WEATHERED BLOCKS

The relation of retention at treatment to block weight loss; Madison data.

Preservative†											
6			7			8			11		
n	R*	Weight loss per cent	n	R	Weight loss per cent	n	R	Weight loss per cent	n	R	Weight loss per cent
<i>Test fungus, Lentinus lepideus, Mad. 534</i>											
6	17.3	1.83	6	17.0	1.77	6	17.4	1.60	6	15.2	2.22
6	14.4	1.94	6	14.1	1.82	6	14.7	1.58	6	14.0	2.22
6	11.5	1.62	6	11.7	1.73	6	11.9	1.57	6	11.7	2.17
6	8.9	2.33	6	8.6	2.92	6	8.9	3.31	6	9.0	1.80
7	6.5	5.73	6	6.2	8.96	6	6.1	7.18	6	6.2	2.57
5	4.3	9.89	6	4.4	11.73	6	4.6	10.61	6	4.2	9.93
6	3.1	13.07	7	3.1	15.77	6	3.1	14.95	6	3.1	15.79
6	2.3	15.44	5	2.3	19.19	6	2.3	16.28	6	2.3	16.84

Test fungus, Lentinus lepideus, Mad. 534

Preservative								
A			B			E		
n	R	Weight loss per cent	n	R	Weight loss per cent	n	R	Weight loss per cent
2	11.70	2.70	2	12.25	3.10	2	11.15	2.55
1	10.10	2.90	2	10.70	2.40	1	9.50	2.60
4	7.20	3.94	3	7.80	2.63	4	8.38	2.55
3	5.43	7.96	4	4.88	2.71	4	5.53	2.56
4	3.60	15.20	3	2.97	3.10	3	3.70	2.63
10	2.15	22.18	3	2.13	3.49	3	2.76	6.04
6	1.00	23.84	4	1.55	6.69	4	2.05	9.49
5	.58	17.35	3	.97	12.34	4	1.50	16.53
			3	.70	25.53	10	.77	29.70
			8	.40	35.43			

TABLE VIII—*Continued*
Test fungus, Lenzites trabea, Mad. 617

Preservative								
A			B			E		
3	9.14	2.57	2	12.35	2.80	3	10.93	2.70
4	7.58	2.42	2	10.30	2.60	3	9.17	2.40
3	5.53	2.73	3	7.80	2.80	2	6.95	2.60
2	3.70	2.81	4	4.95	3.24	3	5.30	2.49
3	3.20	3.97	3	3.33	14.21	4	3.95	3.14
6	2.17	8.57	11	1.51	35.76	3	2.47	8.00
7	1.14	38.96	10	.46	46.63	4	1.85	10.79
7	.60	55.44				6	1.08	52.90
						7	.64	55.06

* R = retention at treatment in lb/cu ft.

† Preservatives 6, 7, 8 and 11 are the numbered coop. creosotes (12).

A = BTL 5340 creosote,

B = 5 per cent penta in petroleum,

E = 50/50 by volume mixture of A and B.

All preservatives were applied in a toluene solution. The heavy lines represent approximate threshold levels.

carefully following the soil-block technique. The data, representing the writer's interpretation of the relation between average weight loss and average treatment retention, are shown in Table VIII and represented by the graphs in Figs. 10, 11 and 12. The preservatives in Figs. 11 and 12 labeled (A), (B) and (E) on the graphs, were respectively a domestic creosote (BTL No. 5340, Table II); a 5 per cent solution of pentachlorophenol in Standard Oil Company of New Jersey No. 2105 Process oil, and a 50/50 by volume mixture of these two.⁴¹ Fig. 11 represents the results obtained with cooperative creosotes Nos. 6, 7, 8 and 11³⁹ and with BTL No. 5340.

In all three figures there are weight losses that can evidently be classed as operational losses, that is, losses by evaporation of some of the volatile materials still remaining in the blocks during the time they were in test and in the subsequent conditioning period.⁴¹ The general areas in which the amount of preservative with which the blocks were treated failed to protect the wood against attack by the different fungi are shown by the rise in the weight loss lines. Perhaps the most interesting set of comparative results are revealed by Fig. 10. The test fungus was *Lentinus lepideus*, Mad. 534. From these graphs the threshold for creosote for this organism

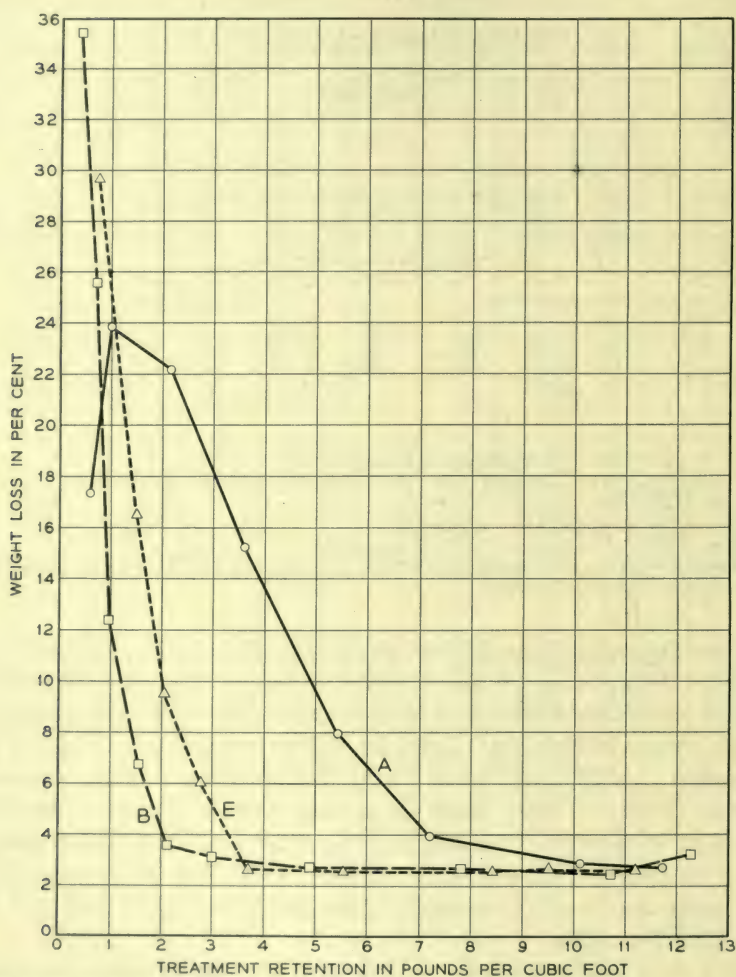


Fig. 10—Soil-block tests against *Lentinus lepideus*; weathered southern pine sapwood blocks; comparison of (A) creosote No. 5340, (B) a 5 per cent solution of pentachlorophenol in Standard Oil Company of New Jersey No. 2105 Process Oil and (E) a 50/50 by volume blend of the two; the relation of operational weight losses, losses by decay, and retention at treatment, lb/cu ft; based on Madison data. The Madison treatment thresholds for these 3 preservative solutions were set at 7.5, 2.4 and 3.4 lb/cu ft, respectively. See text, Table VIII, companion Figs. 11 & 12, Bibliography, Reference 41.

appears to be somewhat in excess of 7.5 pounds per cubic foot, whereas the thresholds for the pentachlorophenol solution are somewhere between 2 and 3 pounds, and the threshold for the mixtures of the creosote and the penta solution about 3.7 pounds per cubic foot. Duncan⁴¹ gives these respective thresholds as 7.5, 2.4 and 3.4 pounds per cubic foot.

Fig. 12 shows that when decay does occur as a result of attack by *Lenzites trabea*, Mad. 617, the loss of weight in the wood is considerably greater than in the case of attack by *Lentinus lepideus*. In the case of *Lenzites trabea*, creosote appears as the best of the three preserva-

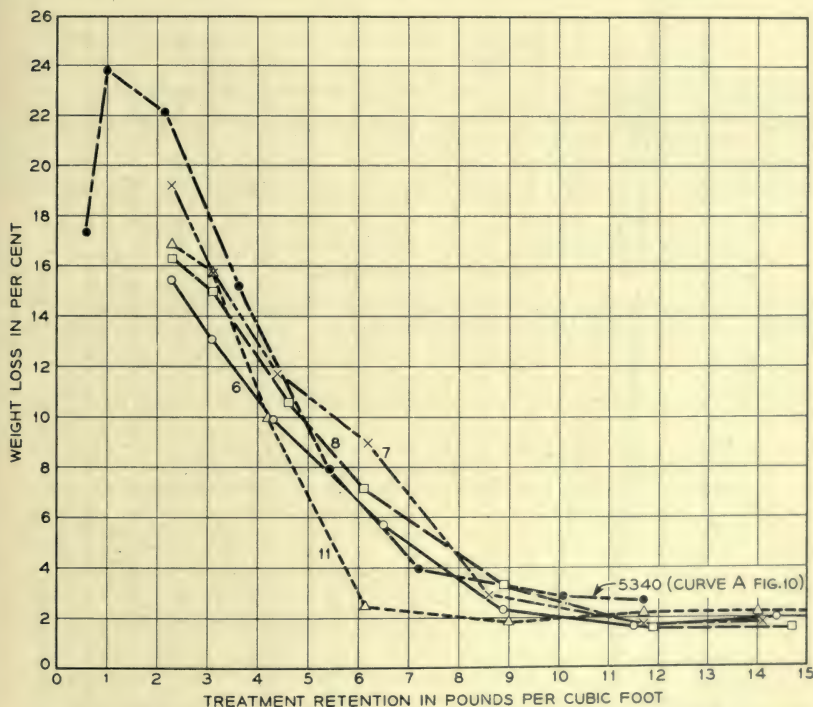


Fig. 11—Soil-block tests against *Lentinus lepideus*; weathered blocks; comparison of cooperative creosotes Nos. 6, 7, 8 and 11, and BTL No. 5340; based on Madison data. The Madison treatment thresholds for these five creosotes were set at 9.0, 9.0, 9.4, 6.5 and 7.5 lb/cu ft, respectively. See Table VIII and Bibliography, References 39 and 41.



Fig. 12—Soil-block tests against *Lenzites trabea*; weathered blocks; comparison of preservatives A, B and E (Fig. 10); based on Madison data. The Madison treatment thresholds for these 3 preservatives were set at 3.2, 4.8 and 4.0 lb/cu ft, respectively. In the poorly protected blocks note the higher per cent weight losses caused by *Lenzites trabea* in comparison with weight losses caused by *Lentinus lepideus* (Figs. 10 & 11). See Table VIII and Bibliography, Reference 41.

tives; and the mixture of creosote and penta solution is somewhat better than the penta solution alone, although the difference is not great.

The results obtained in testing the different creosotes represented in Fig. 11 are similar in general character for the four domestic oils, but oil No. 11, the mixture of British vertical retort tar creosote and British coke oven tar creosote, appears to behave differently. The thresholds for all of these creosotes, as determined by the Madison investigators, are shown in Table XXXV. The figures in this table correspond very closely to thresholds determined by visual observation of the test blocks.

Check Tests at the Murray Hill Laboratories

It will be noted that in Fig. 11 the points used for locating the graphs are rather far apart in the general region of the estimated thresholds. The values shown in Table XXXV were obtained at Madison by the intersection of regression lines drawn through the points representing

TABLE XXXV — SUMMARY AND INTERPRETATION OF
SOIL-BLOCK TESTS

Weathered, creosoted southern pine sapwood blocks; creosote losses; amounts and gross characteristics of residual oils at threshold retentions for *Lentinus lepidus*.

1	2	3	4	5	6	7	8	9	10	11
Item	Creosote No.*	Specific gravity 38/ 15.5°C	Residue above 355°C per cent	Thresh- hold lb/cu ft	Per cent loss	Creosote loss lb/cu ft	Residual creosote lb/cu ft	Calculated residue above 355°C	Residual creosote	
									>355°C lb/cu ft	<355°C lb/cu ft
1	1	1.065	18.5	9.8	53.1	5.2	4.6	39.4	1.81	2.79
2	7	1.077	20.5	9.0	47.8	4.3	4.7	39.3	1.85	2.85
3	2	1.081	30.6	10.2	47.1	4.8	5.4	57.8	3.12	2.28
4	6	1.093	34.2	9.0	37.8	3.4	5.6	55.0	3.08	2.52
5	3	1.108	50.4	12.2	30.3	3.7	8.5	72.3	6.15	2.35
6	8	1.115	53.2	9.4	25.5	2.4	7.0	71.4	5.00	2.00
7	9a		21.2	5.7	40.4	2.3	3.4	35.6	1.21	2.19
8	9	1.001	20.0	5.8	43.1	2.5	3.3	35.1	1.16	2.14
9	10a		14.4	6.7	50.9	3.4	3.3	29.4	0.97	2.23
10	10	1.068	15.2	6.9	52.2	3.6	3.3	31.8	1.05	2.25
11	11	1.038	18.0	6.5	47.7	3.1	3.4	34.4	1.17	2.23
12	M1	1.107	41.9	8.0	33.8	2.7	5.3	63.3	3.35	1.95
13	M2	1.070	18.1	8.3	50.6	4.2	4.1	36.6	1.51	2.59
14	BTL 5340	1.088	20.9	7.5	46.6	3.5	4.0	39.1	1.57	2.43

* Creosotes 1, 2, 3, 6, 7, 8, 9, 10 and 11 are those in use in the Cooperative Creosote Tests (see Bibliography, References 12 and 39. Oils 9a and 10a are samples from the same lots as numbers 9 and 10. (See Bibliography, Reference 36.) For oils M1 and M2 see Bibliography, References 37 and 38. Creosote 5340 is shown in Table II.

operational losses and through the points representing weight losses. Data from a repetition of these tests is desirable in order to establish the thresholds more definitely from actual weight loss or observational data taken close to the assumed threshold points. At Bell Telephone Laboratories a check series of tests is now under way on cooperative creosotes 6, 7 and 8, domestic oils, and creosotes 9, 10 and 11, British oils; and comparison tests are also being run on creosote BTL-5340 and on 5 per cent pentachlorophenol in the 2105 process oil. The aim has been to treat the blocks to a series of retentions that vary narrowly around the thresholds set by the Madison investigators.

Across the Threshold

Fig. 13 is an illustration of representative blocks from the creosote series, line A (creosote BTL-5340) in Fig. 10, just at and below the threshold. Fig. 14 shows the character of the attack by *Lenzites trabea* on blocks treated with a 4.92 per cent solution of pentachlorophenol in Standard Oil Company of New Jersey's 2105 Process Oil in toluene. The blocks are represented at twice their original linear dimensions. The exact nature of the decay is difficult to show. The experimenter has to learn a system of diagnosis that involves both visual observation and the "feel" of the blocks for distortion and firmness that supplement weight loss data. For example, in Fig. 14, a threshold between 0.20 and 0.25 pound of penta per cubic foot (Blocks C and D) is indicated and this conforms closely to the results with the same penta-petroleum solution at Madison.⁴¹

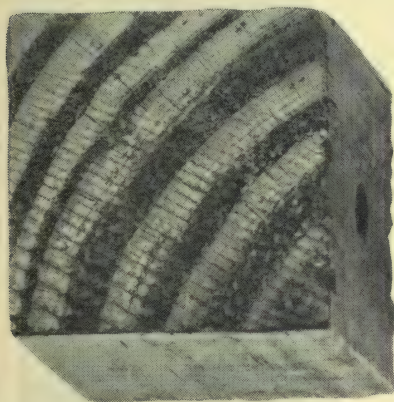
The Significance of the Results of Laboratory Soil-Block Tests on Oil-Type Preservatives

The main conclusions from this discussion of the results of soil-block tests on weathered creosoted wood conducted at Madison are (a) that in general, under the test conditions, at least 8 and sometimes 9 pounds or more of creosote per cubic foot is a necessary treatment to prevent attack by *Lentinus lepideus* on $\frac{3}{4}$ -inch cube blocks of southern pine sapwood; and (b) that a penta petroleum solution is much more effective than creosote against this same organism. As will be emphasized later, this general conclusion about *Lentinus lepideus* and creosote corresponds with the conclusions to be drawn from the interpretation of results of the small stake tests and from the test of pole-diameter posts in the Gulfport test plot.

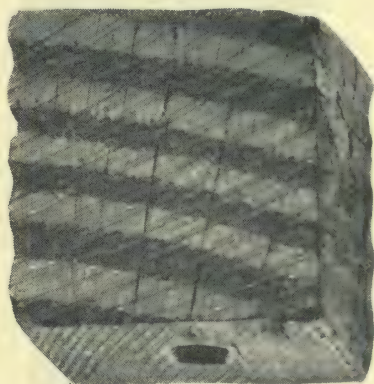
The creosote tested is a better preservative against *Lenzites trabea* than the penta-petroleum, but the creosote threshold for this organism

is below what one would have to use commercially in order to insure protection against *Lentinus lepideus* and to preserve wood exposed to ground contact.

As far as *Lentinus lepideus* is concerned, the best overall preservative combination from the laboratory test would appear to be a 50/50 by volume blend of the creosote and the penta-petroleum solution, since such a blend appears to contain the best in both components. However, there are certain practical considerations, principally relating to the incompatibility of some creosotes and petroleum, which make the combined preservative a difficult one to operate with commercially.



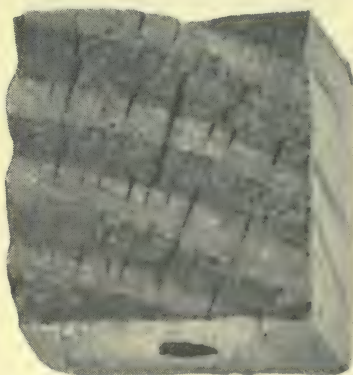
(a)



(b)



(c)

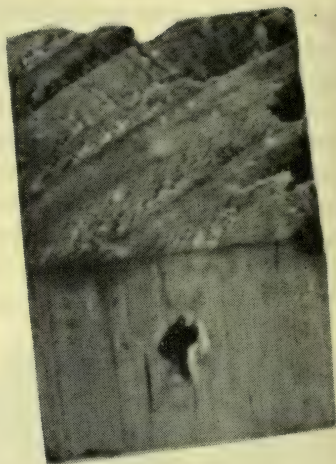


(d)

Fig. 13—Across the threshold—creosote. Test fungus, *Lentinus lepideus*; creosote concentration at treatment: (a) 11.70; (b) 6.92; (c) 5.45; and (d) 4.15 lb/cu ft.



(a)



(b)



(c)



(d)

Fig. 14—Across the threshold—pentachlorophenol solution. Test fungus *Lenzites trabea*; dry penta concentration at treatment: (a) 0.052, (b) 0.099, (c) 0.193; and (d) 0.238 lb/cu ft. Photo by A. H. Hearn.

Relatively speaking the soil-block test procedure is much more rapid than the test plot experiments that are to be discussed next, but since the inferences with respect to retention requirements for creosote appear to be the same for the laboratory and the field tests, the former have a direct and immediate application in practical pole preservation.

BIBLIOGRAPHY

1. Alleman, G., Quantity and Character of Creosote in Well-Preserved Timbers. Circ. 98, Forest Service, U. S. Dept. Agric., May 9, 1907.
2. Alliot, H., Méthode d'Essais des Produits Anticryptogamiques. Inst. Nat. du Bois Bull. Techn. 1, 1945.
3. Amadon, C. H., Recent Observations on the Relation between Penetration, Infection and Decay in Creosoted Southern Pine Poles in Line. Am. Wood Preservers' Assoc., Proc., **35**, pp. 187-197, 1939.
4. Baechler, R. H., Relations between the Chemical Constitution and Toxicity of Aliphatic Compounds. Am. Wood Preservers' Assoc., Proc., **43**, pp. 94-111, 1947.
5. Baechler, R. H., The Toxicity of Preservative Oils Before and After Artificial Aging. Am. Wood Preservers' Assoc., Proc., **45**, pp. 90-95, 1949.
6. Bateman, E., Quantity and Quality of Creosote Found in Two Treated Piles After Long Service. Circ. 199, Forest Service, Forest Products Laboratory Series, U. S. Dept. Agric., May 22, 1912.
7. Bateman, E., Coal-Tar and Water-Gas Tar Creosotes: Their Properties and Methods of Testing. U. S. Dept. Agric., Bul. No. 1036, pp. 57-65, Oct. 20, 1922.
8. Bateman, E., A Theory on the Mechanism of the Protection of Wood by Preservatives. Am. Wood Preservers' Assoc., Proc., 1920, p. 251; 1921, p. 506; 1922, p. 70; 1923, p. 136; 1924, p. 33; 1925, p. 22; 1927, p. 41.
9. Bavendam, W., Die pilzwidrige Wirkung der im Holzschutz benutzten Chemikalien. Mitteilungen d. Reichsinst. f. Forst und Holzwirtschaft. No. 7, 20 pp. Aug., 1948.
10. Baxter, D. V., *Pathology in Forest Practice*. 2nd Ed. XI + 601 pp. John Wiley and Sons, New York, 1952.
11. Bertleff, V., Prüfung der Fungiziden Eigenschaften Ostrauer Steinkohlenteer Imprägnierölle und ihrer Bestandteile. Chem. Zeitung, **63**, p. 438, 1939.
12. Bescher, R. H., and others, Cooperative Creosote Tests. Am. Wood Preservers' Assoc., Proc., **46**, pp. 68-79, 1950.
13. Bienfait, J. L., and T. Hof, Buitenproeven met geconserveerde palen. I ste mededeling. Circ. 8 Ser. III, Conservering en Veredeling No. 3, Central Instituut voor Materiaal Onderzoek. Delft, Holland, Dec., 1948.
14. Bland, D. E., A Study of Toxicity of Australian Vertical Retort Creosote Oils to *Lentinus Lepideus* Fr., *Polystictus Versicolor* (L) Fr., and Madison 517. Australian Council Sci. and Ind. Jl. **15**, pp. 135-146, 1942.
15. Blew, J. O., Comparison of Wood Preservatives in Stake Tests. Am. Wood Preservers' Assoc., Proc., **44**, pp. 88-119, 1948.
16. Blew, J. O., C. A. Richards, and R. H. Baechler, Evaluating Wood Preservatives. For. Prod. Res. Society, Proc., **5**, pp. 230-238, (Wood-block tests, pp. 235-236), 1951.
17. Blew, J. O., Comparison of Wood Preservatives in Mississippi Post Study (1952 Progress Report). Report No. R1757, Forest Products Laboratory, U. S. Forest Service, Madison, Wis.
18. Blew, J. O., Comparison of Wood Preservatives in Stake Tests (1952 Progress Report). Report No. D1761, Forest Products Laboratory, U. S. Forest Service, Madison, Wis.
19. Boyce, J. S., *Forest Pathology*. 550 p. McGraw-Hill Book Company, Inc., New York, 1948.

20. Breazzano, A., Metodo Biologico di Controllo dei Sistemi di Preservazione, dei Legnami Adottato dall' Istituto Sperimentale delle FF. SS. (English translation: Biological Method of Testing Systems of Wood Preservation Adopted by the State Railway Experiment Institute.) Tech. Rev. Italian Railways, 4, No. 5, pp. 3-8, Nov., 1913. Forest Products Laboratory, Madison, Wis.
21. Breazzano, A., Metodi Normali di Prova sulla Putrescibilit  dei Legnami (English translation: Standard Methods of Testing the Putrescibility of Wood.) Extract from Report of the Ninth Meeting of the Italian Association for the Study of Building Materials; Turin; April, 1922. Forest Products Laboratory, Madison, Wis.
22. Breazzano, A., Osservazioni sul Metodo dei Blocchetti di Legno in Usa nell' Analisi Tossimetrica delle Sostanze Conservatrici del Legno. Revista Technica Delle Ferrovie Italiane. 47, No. 6, June 15, 1935. (Observations on the Wood Block Method in the Toximetric Analysis of Wood Preservatives. Excerpt from Technical Review of Italian Railways. English translation; Forest Products Laboratory.)
23. Breazzano, A., Methods for Determining the Fungicidal Power of Wood Preservatives. International Assn. for Testing Materials, Proc., London Congress, Group C—Organic Materials, Sub. group 3, pp. 484-486, April, 1937.
24. British Standard Method of Test for the Toxicity of Wood Preservatives to Fungi. British Standards Institution. British Standard No. 838, 17 pp., April, 1939.
25. Broekhuizen, S., Onderzoekingen over de Conserverende Waarde van een Aantal Houtconserveermiddelen. Rapp. Comm. Gebruikw. inh. hout. Deel II, Bijlage II, pp. 89-122, The Hague, 1937.
26. Cartwright, K. St. G., and W. P. K. Findlay, *Decay of Timber and its Prevention*. VI plus 294 p. London: His Majesty's Stationery Office, 1946. Reprinted 1948.
27. Colley, R. H., The Effect of Incipient Decay on the Mechanical Properties of Airplane Timber. (Abstract) Phytopathology, 11, p. 45, 1921.
28. COLLEY, R. H., T. R. C. WILSON, and R. F. LUXFORD, The Effect of *Polyporus Schweinitzii* and *Trametes Pini* on the Shock-resistance, Compression Parallel to Grain Strength, and Specific Gravity of Sitka Spruce. Forest Products Laboratory, Project L-243-J1, Typewritten Report, 29 p., 32 figs., 4 plates. July 3, 1925.
29. Colley, R. H., and C. H. Amadon, Relation between Penetration and Decay in Creosoted Southern Pine Poles. Bell Sys. Tech. J., 15, pp. 363-379, July, 1936.
30. Colley, R. H., Some Observations on the Selection and Use of Modern Wood Preservatives. Reports, Twenty-fourth Session, Communications Section, Association of American Railroads, October, 1947, pp. 17-25.
31. Colley, R. H., Wood preservation and Timber Economy. Forest Products Institute of Canada. Papers presented at the First Annual Convention, Ottawa, Oct. 30-31, 1950.
32. Curtin, L. P., B. L. Kline, and W. Thordarson, V—Weathering Tests on Treated Wood. Ind. and Eng. Chem. 10, No. 12, pp. 1340-1343, Dec., 1927.
33. DIN (Deutsche Normen) DVM 2176, Blatt 1. Pr fung von Holzschutzmitteln. Mykologische Kurzpr fung (Kl tzen Verfahren). Berlin, Aug. 1939. (New Edition DIN 52176).
34. DIN DVM 52176, Blatt 2. Pr fung von Holzschutzmitteln. Bestimmung der Auslaugbarkeit. Berlin, May 1941. (Reprinted Oct., 1948).
35. Duncan, C. G., and C. A. Richards, Methods of Evaluating Wood-Preservatives: Weathered Impregnated Wood Blocks. Am. Wood Preservers' Assoc., Proc., 44, pp. 259-264, 1948.
36. Duncan, C. G., A comparison of Two English Creosotes Produced from Coke-oven Coal Tar and Vertical-retort Coal Tar. Mss. Office Report, Division of Forest Pathology, Bur. Pl. Ind., Forest Prod. Lab., Madison, Wis., Jan. 21, 1949.
37. Duncan, C. G., and C. A. Richards, Evaluating Wood Preservatives by Soil-block Tests: 1. Effect of Carrier on Pentachlorophenol Solutions; 2. Com-

- parison of a Coal Tar Creosote, a Petroleum Containing Pentachlorophenol or Copper Naphthenate and Mixtures of Them. Amer. Wood Preservers Assoc., Proc., **46**, pp. 131-145, 1950.
38. Duncan, C. G., and C. A. Richards, Evaluating Wood Preservatives by Soil-Block Tests: 3. The Effect of Mixing a Coal Tar Creosote and a Pentachlorophenol Solution with a Petroleum; a Creosote with a Coke Oven Tar or Pentachlorophenol Solution. Amer. Wood Preservers' Assoc., Proc., **47**, pp. 264-274, 1951.
 39. Duncan, C. G., and C. A. Richards, Evaluating Wood Preservatives by Soil-Block Tests: 4. Creosotes. Amer. Wood Preservers' Assoc., Proc., **47**, pp. 275-287, 1951.
 40. Duncan, C. G., Evaluating Wood Preservatives by Soil-Block Tests: 5. Lignite-Tar and Oil-Tar Creosotes. Amer. Wood Preservers' Assoc., Proc., **48**, 1952.
 41. Duncan, C. G., Soil-Block versus Agar-Block Techniques for Evaluation of Oil-Type Preservatives: Creosote, Copper Naphthenate and Pentachlorophenol. Mss. Division of Forest Pathology, Bur. Pl. Ind., Forest Products Lab; Madison, Wis., April 29, 1952.
 42. Eden, Johan, och Erik Rennerfelt, Undersöknar enligt klotsmetoden av några fräimpregningsmedel. (Studies on Wood Preservatives, According to the Block Method) Meddelanden från Statens Skogsforskningsinstitut, Bd. **35**, No. 10, 1946.
 43. Eden, Johan, och Erik Rennerfelt, Fält och rökammarförsök avsedda att utröna skyddsverkan hos olika fräimpregningsmedel. (Field and Decay-chamber Experiments to Ascertain the Protective Effect of Various Wood Preservatives.) Meddelanden från Statens Skogsforskningsinstitut. Bd. **38**, No. 4, 1949.
 44. Falck, R., Die wichtigsten reinen Holzschutzmittel, die Methoden ihre Wertzahlen, Eigenschaften und Anwendung. Hausschwammforschungen, **8**, pp. 18-20, 1927.
 45. Findlay, W. P. K., A Standard Laboratory Test for Wood Preservatives. British Wood Preserving Assoc., Jl., **5**, pp. 89-93, 1935.
 46. Finholt, R. W., Improved Toximetric Agar-dish Test for Evaluation of Wood Preservatives. Anal. Chem., **23**, No. 7, pp. 1039-1039, July, 1951.
 47. Finholt, R. W., M. Weeks, and C. Hathaway, New Theory on Wood Preservation. Ind. and Eng. Chem. **44**, No. 1, pp. 101-105, Jan., 1952.
 48. Flerov, B. C., und C. A. Popov., Methode zur Untersuchung der Wirkung von antiseptische Mitteln auf holzerstörende Pilze. Angew. Bot. **15**, pp. 386-406, 1933.
 49. Froesch, C. J., V—The Correlation of Distillation Range with Viscosity of Creosote. Physics, **6**, pp. 165-170, May, 1935.
 50. Gillander, H. E., C. G. King, E. O. Rhodes, and J. N. Roche, The Weathering of Creosote. Ind. and Eng. Chem., **26**, No. 2, pp. 175-183, Feb., 1934.
 51. Harrow, K. M., Toxicity of Water-Soluble Wood-Preservatives to Wood-Destroying Fungi. New Zealand Jl., Sec. B., **31**, No. 5, pp. 14-19, Mar., 1950.
 52. Harrow, K. M., Note on the Soil Moisture Content Used with the Leutritz Technique for Testing Toxicity of Wood Preservatives Against Fungi. New Zealand Jl. Sci. and Tech., **4**, pp. 39-40, Jan., 1951.
 53. Hatfield, I., Information on Pentachlorophenol as a Wood Preserving Chemical. Am. Wood Preservers' Assoc., Proc., **40**, pp. 47-65, 1944.
 54. Holzschutzmittel Prüfung und Forschung III. Wissenschaftliche Abhandlungen der Deutschen Materialprüfungsanstalten: Berlin-Dahlem, Heidelberg, 1950.
 1. Schulze, B., G. Theden u K. Starfinger, Ergebnisse einer vergleichenden Prüfung der Pilzwidrigen Wirksamkeit von Holzschutzmitteln. pp. 1-40.
 2. Becker, G., Ergebnisse einer vergleichenden Prüfung der insekttötenden Wirkung von Holzschutzmitteln. II Teil. pp. 40-62.
 3. Becker, G., Prüfung der "Tropeneignung" von Holzschutzmitteln gegen Termiten. pp. 62-76.
 4. Becker, G. u B. Schulze, Laboratoriumsprüfung von Holzschutzmitteln gegen Meerwasserschädlinge. pp. 76-83.

55. Hopkins, C. Y., and B. B. Coldwell, Surface Coatings for Rotproofing Wood. Canadian Chemistry and Process Industries. N. R. C. No. 1256, Dec., 1944.
56. Howe, P. J., Weathering and Field Tests on Treated Wood. Amer. Wood Preservers' Assoc., Proc., 1928, p. 192.
57. Hubert, E. E., A study of Laboratory Methods Used in Testing the Relative Resistance of Wood to Decay. Univ. of Idaho Bulletin, **34**, No. 15, July 1929.
58. Hubert, E. E., *An Outline of Forest Pathology*. 543 pp. John Wiley and Sons, New York, 1931.
59. Hudson, M. S., and R. H. Baechler, The Oxidation of Creosote—Its Significance in Timber Treating Operations. Amer. Wood Preservers' Assoc., Proc., **36**, pp. 74-112, 1940.
60. Hudson, M. S., *Poria radiculosa*, a Creosote Tolerant Organism. For. Prod. Res. Soc., **11**, 2, No. 2, pp. 73-74, June, 1952.
61. Humphrey, C. J., and R. M. Flemming, The Toxicity to Fungi of Various Oils and Salts, Particularly Those Used in Wood Preservation. Bul. No. 227, U. S. Dept. Agric., Aug. 23, 1915.
62. Hunt, G. M., and G. A. Garratt, *Wood Preservation*. VIII, 457 p. McGraw-Hill, New York, 1938. (In course of revision.)
63. Hunt, G. M., and T. E. Snyder, An International Termite Exposure Test—Twenty-First Progress Report. Amer. Wood Preservers' Assoc., Proc., **48**, 1952.
64. Jacquot, C., Controle de l'Efficacité des Fongicides Utilisés pour l'Imprégnation des Bois. Étude Critique de la Technique Standard Anglaise et de la Norme Allemande DIN DVM 2176. Principes pour l'Établissement d'une Norme Française. Extr. d'Ann. l'Ecole Eaux Forêts, **8**, pp. 185-206, 1942.
65. Kaufert, F. H., A Survey of Laboratory Methods Used in the Evaluation of Wood Preservatives. Report of Committee P-6, Appendix A. Amer. Wood Preservers' Assoc., Proc., **45**, pp. 55-59, 1949.
66. Leutritz, J., Jr., The Toxic Action of Various Compounds on The Fungus *Lentinus Lepideus* Fr.). Unpublished Thesis, Columbia University, Nov., 1933.
67. Leutritz, J., Jr., Laboratory Tests of Wood Preservatives. Bell Lab. Record, **16**, No. 9, pp. 324-328, May, 1938.
68. Leutritz, J., Jr., Acceleration of Toximetric Tests of Wood Preservatives by the Use of Soil as a Medium. Phytopathology, **39**, No. 10, pp. 901-903, Oct., 1939.
69. Leutritz, J., Jr., Outdoor Tests of Wood Preservatives. Bell Lab. Record, **22**, No. 4, pp. 179-182, Dec., 1943.
70. Leutritz, J., Jr., A Wood-Soil Contact Culture Technique for Laboratory Study of Wood-Destroying Fungi, Wood Decay and Wood Preservation. Bell Sys. Tech. J., **25**, No. 1, pp. 102-135, Jan., 1946.
71. Liese, Nowak, Peters, Rabanus, Krieg, and Pflug, Toximetrische Bestimmung von Holzkonservierungsmitteln. Angew. Botanik, pp. 484-488, Nov.-Dec., 1935.
72. Liese, J. et al., Toximetrische Bestimmung von Holzkonservierungsmitteln. Angew. Chemie, **48**, Beihefte 11, 1935.
73. Loseby, P. J. A., and P. M. D. Krog, The Persistence and Termite Resistance of Creosote and Its Constituent Fractions. Jour. South African Forestry Assoc., **11**, No. 11, pp. 26-32, June, 1944.
74. Lumsden, G. Q., and A. H. Hearn, Greensalt Treatment of Poles. Amer. Wood Preservers' Assoc., Proc., **38**, pp. 349-361, 1942.
75. Lumsden, G. Q., Proving Grounds for Telephone Poles. Bell Lab. Record, **22**, pp. 12-14, Sept., 1943.
76. Lumsden, G. Q., A Quarter Century of Evaluation of Wood Preservatives in Poles and Posts at the Gulfport Test Plot. Amer. Wood Preservers' Assoc., Proc., **48**, 1952.
77. Lutz, M. L., Méthodes Permittant de Déterminer la Résistivité des Bois Bruts ou Immunisés Soumis a l'Attaque par les Champignons Lignicoles. Ann. l'Ecole Nat. Eaux Forêts **5**, pp. 317-327, 1935.
78. Mahlke-Troschel-Liese, Holzkonservierung (Wood Preservation), 3rd Ed. XII + 571 p. Springer-Verlag, Berlin/Göttingen/Heidelberg, 1950.

79. McMahon, W., C. M. Hill, and F. C. Koch, Greensalt—A New Preservative for Wood. Amer. Wood Preservers' Assoc., Proc., **38**, pp. 334-348, 1942.
80. Martin, S. W., Characterization of Creosote Oils. Amer. Wood Preservers' Assoc., Proc., **45**, pp. 100-130, 1949.
81. Mayfield, P. B., The Toxic Elements of High Temperature Coal Tar Creosote. Amer. Wood Preservers' Assoc., Proc., **47**, pp. 62-85, 1951.
82. Narayanamurti, D., V. Ranganathan, Ragbir Singh, T. R. Chandrasekhar, and A. Banerjee, Studies on Coal Tar Creosote as a Wood Preservative, Part II. Indian Forest Bulletin, No. 144, 1948, 7 + 43 pp., JI. of India Press, Calcutta, 1950.
83. Peters, F., W. Krieg and H. Pflug, Toximetrische Prüfung von Steinkohlen-teeröl. Chem. Zeit., **61**, pp. 275-285, 1937. (English Edition, Pub. Int. Adv. Off. Wood Pres. The Hague. 1937.)
84. Preservative Treatment of Poles. (Condensed from report by American Telephone and Telegraph Company of Aug. 3, 1931.) Am. Wood Preservers' Assoc., Proc., p. 237, 1932.
85. Preservatives Committee; Report of Committee P-6. Am. Wood Preservers' Assoc., Proc., **48**, 1952.
86. Rabanus, Ad., Die Toximetrische Prüfung von Holzkonservierungsmitteln. Angew. Bot. 13, p. 352-371, 1931. (Partial translation in English. Am. Wood Preservers' Assoc., Proc., pp. 34-43, 1933.)
87. Reeve, C. S., Comment on Creosote-Permanence Toxicity Relationships. Am. Wood Preservers' Assoc., Proc., p. 78-79, 1934.
88. Rennerfelt, Erik, och Bo Starkenberg., Träskyddskomittens fält- och rök-kammarförsök. (Field and Decay-Chamber Experiments with Wood Preservatives.) Meddelanden från Statens Skogsforskningsinstitut, Bd., **40**, No. 4, 1951.
89. Rhodes, E. O., J. N. Roche, and H. E. Gillander, Creosote Permanence-Toxicity Relationships. Am. Wood Preservers' Assoc., Proc., pp. 65-78, 1934.
90. Rhodes, E. O., History of Changes in Chemical Composition of Creosote. Am. Wood Preservers' Assoc., Proc., **47**, pp. 40-61, 1951.
91. Rhodes, F. H., and F. T. Gardner, Comparative Efficiencies of the Components of Creosote Oil as Preservatives for Timber. Ind. and Eng. Chem., **22**, No. 2, pp. 167-171, Feb., 1930.
92. Rhodes, F. H., and I. Erickson, Efficiencies of Tar Oil Components as Preservative for Timber. Ind. and Eng. Chem. **25**, pp. 989-991, Sept., 1933.
93. Richards, A. P., Cooperative Creosote Program; Preliminary Progress Report on Marine Exposure Panels. Am. Wood Preservers' Assoc., Proc., **48**, 1952.
94. Richards, C. A., Methods of Testing Relative Toxicity of Wood Preservatives. Am. Wood Preservers' Assoc., Proc., **19**, pp. 127-135, 1923.
95. Richards, C. A., and R. M. Addoms, Laboratory Methods for Evaluating Wood Preservatives: Preliminary Comparison of Agar and Soil Culture Techniques Using Impregnated Wood Blocks. Am. Wood Preservers' Assoc., Proc., **43**, pp. 41-56, 1947.
96. Richards, C. A., Laboratory Decay Resistance Tests—Soil-block Method. (In Cooperative Creosote Tests by R. H. Bescher et al.) Am. Wood Preservers' Assoc., Proc., **46**, pp. 71-76, 1950.
97. Scheffer, T. C., Progressive Effects of Polyporus Versicolor on the Physical and Chemical Properties of Red Gum Sapwood. U. S. Dept. Agric., Tech. Bul., No. 527, Sept., 1936.
98. Scheffer, T. C., T. R. C. Wilson, R. F. Luxford, and Carl Hartley, The Effect of Certain Heart Rot Fungi on the Specific Gravity and Strength of Sitka Spruce and Douglas-Fir. U. S. Dept. Agric., Tech. Bul., No. 779, 24 pp., May, 1941.
99. Schmitz, H., Laboratory Methods of Testing the Toxicity of Wood Preservatives. Ind. and Eng. Chem., Anal. Ed., **1**, No. 7, pp. 76-79, April, 1929.
100. Schmitz, H., and Others A Suggested Toximetric Method for Wood Preservatives. Ind. and Eng. Chem., Anal. Ed., **2**, p. 361, 1930.
101. Schmitz, H., and S. J. Buckman, Toxic Action of Coal-Tar Creosote. Ind. and Eng. Chem., **24**, No. 7, pp. 772-777, 1932.
102. Schmitz, H., W. J. Buckman and H. von Schrenk, Studies of the Biological

- Environment in Treated Wood in Relation to Service Life. Changes in the Character and Amount of 60/40 Creosote-Coal Tar Solution and Coal Tar and Decay Resistance of the Wood of Red Oak Crossties after Five Years Service. *Am. Wood Preservers' Assoc., Proc.*, **37**, pp. 248-297, 1941.
103. Schmitz, H., H. von Schrenk, and A. L. Kammerer, Studies of the Biological Environment in Treated Wood in Relation to Service Life, III. *Am. Wood Preservers' Assoc., Proc.*, **41**, pp. 153-179, 1945.
 104. Schulze, B., und G. Becker, Untersuchungen über die pilzwidrige und insektentötende Wirkung von Fraktionen und Einzelstoffen des Steinkohlenteeröls. *Holzforschung.*, **2**, No. 4, pp. 95-127, 1948.
 105. Sedziak, H. P., The Wood-Block Soil Method of Accelerated Testing of Wood Preservatives. Report of Committee P-6, Appendix B. *Am. Wood Preservers' Assoc., Proc.*, **46**, pp. 55-58, 1950.
 106. Sedziak, H. P., The Evaluation of Two Modern Wood Preservatives. *For. Prod. Res. Soc., Proc.*, 1952.
 107. Snell, W. H., The Use of Wood Discs as a Substrate in Toxicity Tests of Wood Preservatives. *Am. Wood Preservers' Assoc., Proc.*, **25**, pp. 126-129, 1929.
 108. Snell, W. H., and L. B. Shipley, Creosotes—Their Toxicity, Permanence and Permanence of Toxicity. *Am. Wood Preservers' Assoc., Proc.*, **32**, pp. 32-114, 1936.
 109. Standard of N. W. M. A., Method for Testing the Preservative Property of Oil-soluble Wood-Preservatives by Using Wood Specimens Uniformly Impregnated. *Nat. Wood Mfg. Assoc.*, M-1-51, April 27, 1951.
 110. STAS 650-49, Incercarea Toxicității Substanțelor de Impregnat Contra Ciupercilor. *Comisiunea de Standardizare (Rumania)* April 1, 1950.
 111. Suolahti, Osmo, Über Eine das Wachstum von Fäulnispilzen Beschleunigende Chemischen Fernwirkung von Holz. *Statens Tekniska Forskningsanstalt*. 95 p. Helsinki, Finland, 1951.
 112. Tamura, T., New Methods of Test on the Toxicity and Preservative Value of Wood Preservatives. *Phytopathologische Zeitschrift*, **3**, No. 4, pp. 421-437, 1931.
 113. Teesdale, C. H., Volatilization of Various Fractions of Creosote After Their Injection into Wood. *Circ. 188, Forest Service, Forest Products Laboratory Series, U. S. Dept. Agric.*, Oct. 17, 1911.
 114. Tippo, O., J. M. Walter, S. J. Smucker, and W. Spackman, Jr., The Effectiveness of Certain Wood Preservatives in Preventing the Spread of Decay in Wooden Ships. *Lloydia*, **10**, pp. 175-208, Sept., 1947.
 115. Trendelenburg, R., Über die Abkürzung der Zeitdauer von Pilzversuchen an Holz mit Hilfe der Schlagbiegeprüfung. *Holz als Roh- und Werkstoff*. **3**, No. 12, s. 397-407, Dec., 1940.
 116. van den Berge, J. Beoordeeling van de Waarde van Fungicide Stoffen voor Houtconserveering. 183 p. N. V. Technische Boekhandel, J. Waltman, Delft, Holland, 1934.
 117. van Groenou, H. Broese, Weatheringsproeven met Houtconserveermiddelen. (Weathering Tests with Wood Preservatives). *Materiaalencennis*, **7**, No. 10, pp. 63-65, Oct., 1940.
 118. van Groenou, H. Broese, H. W. L. Rischen and J. van den Berge, Wood Preservation During the Last 50 Years. XII + 318 p. A. W. Sijthoff, Leiden, Holland, 1951.
 119. von Pechmann, H., u. O. Schaile, Über die Änderung der dynamischen Festigkeit und der Chemischen Zusammensetzung des Holzes durch den Angriff Holzzerstörender Pilze. *Forstwissenschaftliches Zentralblatt*, **69**, No. 8, s. 441-465, 1950.
 120. Verrall, A. F., Progress Report on Tests of Soak and Brush Preservative Treatments for Use on Wood off the Ground. *Southern Lumberman*, Aug. 1, 1946.
 121. Verrall, A. F., Decay Protection for Exterior Woodwork. *Southern Lumberman*, June 15, 1949.
 122. Warner, R. W., and R. L. Krause, Agar-Block and Soil-Block Methods for Testing Wood Preservatives. *Ind. and Eng. Chem.*, **43**, No. 5, pp. 1102-1107, May, 1951.

123. Waterman, R. E., Forecasting the Behavior of Wood Preservatives. *Bell Lab. Record*, **11**, No. 3, pp. 67-72, Nov., 1932.
124. Waterman, R. E., F. C. Koch, and W. McMahon, Chemical Studies of Wood Preservation. III. Analysis of Preserved Timber. *Ind. and Eng. Chem., Anal. Ed.*, **6**, No. 6, pp. 409-413, Nov. 15, 1934.
125. Waterman, R. E. and R. R. Williams, Chemical Studies of Wood Preservation. IV. Small Sapling Method of Evaluating Wood Preservatives. *Ind. and Eng. Chem., Anal. Ed.*, **6**, No. 6, pp. 413-418, Nov. 15, 1934.
126. Waterman, R. E., J. Leutritz, and C. M. Hill, A Laboratory Evaluation of Wood Preservatives. *Bell System Tech. Jl.*, **16**, pp. 194-211, April, 1937.
127. Waterman, R. E., J. Leutritz and C. M. Hill, Chemical Studies of Wood Preservation: The Wood-block Method of Toxicity Assay. *Ind. and Eng. Chem., Anal. Ed.*, **10**, No. 6, pp. 306-314, June, 1938.
128. Weiss, J. M., The Antiseptic Effect of Creosote Oil and Other Oils Used for Preserving Timber. *Soc. Chem. Ind. Jl.*, **30**, No. 23, Dec. 15, 1911.
129. Williams, R. R., Chemical Studies of Wood Preservation. I. The Problem and Plan of Attack. *Ind. and Eng. Chem., Anal. Ed.*, **6**, No. 5, pp. 308-310, Sept. 15, 1934.

Motion of Gaseous Ions in Strong Electric Fields

By GREGORY H. WANNIER

(Manuscript received August 20, 1952)

This paper applies the Boltzmann method of gaseous kinetics to the problem of charged particles moving through a gas under the influence of a static, uniform electric field. The particle density is assumed to be vanishing low, and the ion-atom collisions are assumed elastic, but the field is taken to be strong; that is the energy which it imparts to the charges is not assumed negligible in comparison to thermal energy. In Part I, the formal framework of such a theory is built up; the motion in the field is describable by the drift velocity concept, and the smoothing out of density variations as an anisotropic diffusion process. In Part II, the "high field" case is treated in detail; this is the case, for which thermal motion of the gas molecules is negligible; the equation is solved completely for the case that the mean free time between collisions may be treated as independent of speed; complete solutions are also presented for extreme mass ratios of the ions and the molecules; special attention is given to the case of equal masses, which has to be handled by numerical methods. In Part III, information about the "intermediate field" case is collected; with the help of a convolution theorem the case of constant mean free time is solved; beyond this, only the case of small ion mass (electrons) is available. In Part IV, the diffusion process, whose existence was proved in Part I, is pushed through to numerical results. Part V discusses the scope of the results achieved and demonstrates the possibility of extending them semiquantitatively beyond their original range.

PART I—GENERAL THEORY OF STRONG FIELD MOTION

IA. QUALITATIVE DISCUSSION

It is well known that if we consider a mixture of gases under no external forces the steady velocity distribution which establishes itself in the mixture does not depend on the interactions between the gas molecules; we have always a Maxwellian distribution for each species

with a temperature common to all. This result arises from statistical mechanics; the derivation of it is simple and requires few assumptions, yet it enjoys a wide degree of generality. As soon, however, as a non-equilibrium feature is imposed upon the system this simplicity vanishes, and the subject acquires ramifications. Results must now be derived by kinetic theory. The amount of labor required increases, while, at the same time, the result achieved becomes less general.

A mixture of charged particles (ions or electrons; in the following often simply referred to as ions) and gas molecules can in principle never be in equilibrium since the presence of the former in itself represents an instability. However, one might expect, that equilibrium exists in a restricted sense, for instance, as regards motion. Even this is rarely the case under actual conditions of observation. The non-equilibrium features of greatest importance for analyzing ion motion are a constant force (electric field) acting upon one species but not the other (mobility theory), and a concentration gradient for one particular species (diffusion theory). It is the purpose of this paper to apply kinetic theory to these problems, and to compute with its help the most important properties which such a gas of charged particles possesses. The work will be distinguished from similar ones in that the electric field will not be supposed weak; velocity distributions which have no resemblance to the Maxwellian distribution will thus make their appearance. Furthermore, the mass of the charged particles will not be assumed small, which means the possibility of getting results for gaseous ions as well as electrons. Magnetic fields, plasma and A.C. phenomena will, however, be excluded. The quantities of interest under those conditions are the drift velocity of the ions, their energy, energy partition and diffusion constants. These quantities will be calculated by assuming plausible mechanical models. The work just outlined has been published in part in abbreviated form in the *Physical Review*;¹ the exposition to follow will, however, proceed independently from these articles.

Much of the work which concerns itself with transport processes in gases makes use of perturbation theory. This method permits us to predict the behavior of a gaseous assembly under an electric field or a concentration gradient in the limit when the field or the gradient are vanishingly small. The result of so perturbing a Maxwellian distribution can be expressed through certain constants, such as the mobility or the diffusion coefficient, which involve the Maxwellian distribution *and* the internal interactions, but not the perturbation itself.

¹ Wannier, G. H., *Phys. Rev.*, **83**, p. 281, 1951 and *Phys. Rev.*, **87**, p. 795, 1952.

The limits of such a procedure can easily be estimated. In the case of an electric field, perturbation techniques apply if the kinetic energy acquired by the ion from the field is small compared to thermal energy. This means at least that the energy acquired in one mean free path be small, i.e.,

$$eE\lambda \ll kT$$

where e is the electronic charge, E the electric field, k Boltzmann's constant, T the absolute temperature, and λ the mean free path. Actually the situation is not even that favorable. If the mass of the ions and the molecules is very different, the energy transferred upon collision is small, and hence the ions possess the ability to store the acquired energy through many collisions; for this reason, the inequality reads more properly

$$\left(\frac{M}{m} + \frac{m}{M}\right) eE\lambda \ll kT,$$

where m is the mass of the ions and M the mass of the gas molecules. After some substitutions this estimate becomes

$$\left(\frac{M}{m} + \frac{m}{M}\right) eE \ll p\sigma, \quad (1)$$

where p is the true gas pressure and σ the collision cross-section. Taking as an example an ion travelling in the parent gas we find

$$\frac{E}{p} \ll 2 \frac{\sigma}{e} \sim 2 \cdot \frac{4\pi \cdot 10^{-16}}{5 \cdot 10^{-10}} = 5 \cdot 10^{-6} \text{ e.s.u.}$$

or in commonly employed units

$$\frac{E}{p} \ll 2 \text{ volt/cm (mm Hg)}.$$

It is clear that this limit is often surpassed in experimental situations.

The cases in which the limit (1) is applicable are of no further interest here because they are well covered in the literature.² A field will be called "low" when it satisfies the criterion (1) and "high" when the inequality is reversed. It is important to notice that a fixed field at a fixed gas density may shift from "low" to "high" through a drop in temperature.

All calculations to follow will contain the assumption of "low ion concentration" which is often made in studies of this sort. It means that

² See for instance: A. M. Tyndall, *The Mobility of Positive Ions in Gases*, Cambridge University Press, 1938, Chapter IV.

all effects which ions exert upon each other are neglected. The equation for the distribution function of ionic velocities is then linear instead of quadratic. It is clear that this simplification presents great advantages from the point of view of calculation.

In deriving a criterion for the validity of this assumption we must distinguish two types of effects of the ions upon each other. The first is the space charge effect. In this effect the ions at large distances make the major contribution. Its magnitude depends on apparatus dimensions. The criterion for no space charge distortion of the field E is

$$n \ll \frac{E}{4\pi eX} \quad (2)$$

where n is the number density of the ions and X a suitable length chosen from apparatus dimensions. Inequality (2) is quite stringent because it predicts field distortions at values of n of the order of 10^8 cm^{-3} . This is the value at which it will become impossible, or at least difficult, to make significant experimental measurements. But from the point of view of theory this criterion is not relevant. Space charge does not change the character of the velocity distribution of the ions because the type of ion-ion interaction producing the space charge field is long range and creates only a smooth modification of the electric field which we may presume to have been included in the original field. What we are concerned with here are ion-ion interactions which have a random character and thus are apt to upset a velocity distribution derived from the "low concentration" theory. From this point of view neighboring ions are most effective because their relative location fluctuates rapidly, and hence, the Coulomb force between them will induce mutual scattering. The magnitude of this force is of the order $e^2 n^{2/3}$ where n is the number density of the ions. It is known from theory³ that the effect of a Coulomb force is preferably not represented by discrete "collisions" but by a continuous bending of the entire path. Thus we come to the conclusion that random ion-ion forces have no effect if the force given above cannot produce a significant deflection in one mean free path. This means

$$e^2 n^{2/3} \lambda \ll \text{mean ion energy} \quad (3)$$

According to whether we are in the high or low field region we get different criteria from this. At low field the thermal energy predominates and we get

$$e^2 n^{2/3} \ll p\sigma \quad (3a)$$

³ Mott and Massey, *The Theory of Atomic Collision*, Oxford Press 1933, Chapter III.

At high field the "field" energy predominates and we get

$$e^2 n^{2/3} \ll eE \left(\frac{M}{m} + \frac{m}{M} \right) \quad (3b)$$

A rough evaluation of inequality (3a) for one mm Hg pressure gives

$$n^{2/3} \ll \frac{10^3 \cdot 4\pi \cdot 10^{-16}}{25 \cdot 10^{-20}} = \frac{1}{2} \cdot 10^7 \text{ cm}^{-2}$$

$$n \ll 10^{10} \text{ particles/cm}^3$$

This corresponds to a current of about 10^{15} particles/cm² sec or 200 μ amps/cm². At lower pressure the criterion becomes more stringent. Equation (3b) gives similar results.

It is appropriate to survey at this point the past theoretical work treating the "low concentration" theory of ionic motion for arbitrary fields. A rather complete body of work exists for electrons where the following three assumptions seem appropriate: (a) that the mass of an "ion" is very small compared to the mass of a molecule, (b) that the total kinetic energy is conserved in each encounter, and (c) that the angular distribution is isotropic in the center of mass system.

These three assumptions lead to a distribution law given by Chapman and Cowling.⁴ The law has considerable flexibility because it permits the substitution of an arbitrary relationship connecting mean free path and speed of encounter. In addition it contains no assumption as to whether we have low or high field. A more specialized and explicit distribution law is obtained if we assume in addition: (d) that the collision cross-section is independent of the speed of encounter (hard sphere approximation); and (e) that we deal with the high field case only. The special law resulting in this case is the distribution law of Druyvesteyn.

If an improvement over the Chapman-Cowling distribution for electrons is desired account should be taken of inelastic collisions, that is assumption (b) should be discarded. Work in that direction has been carried out by Smit, Allen⁵ and others.

The assumption to be discarded first in theory of ionic motion is, of course, assumption (a). In order to understand what this implies we must understand what advantages assumption (a) has in a calculation. In the limit when the ionic mass is very small the encounters with gas

⁴ Chapman-Cowling, *The Mathematical Theory of Non-uniform Gases*, Cambridge University Press 1939, Sections 18.7-18.74. Other references are found there.

⁵ Smit, J. A., *Physica*, **3**, p. 543, 1937 and H. W. Allen, *Phys. Rev.*, **50**, p. 707, 1937.

molecules become such that momentum is lost quickly, but energy is accumulated in the form of random motion. As a result of this we end up with a distribution function which is very nearly spherically symmetrical in velocity space. Such a situation permits obvious procedures through which the entire calculation is simplified. These procedures will not longer be available when assumption (a) is dropped.

Knowledge concerning the structure of the velocity distribution function for gaseous ions is practically nonexistent at this time. Hershey, who deals with the motion of ions in the high field case, simply substitutes for it a Maxwellian distribution with an unknown offset of the origin and unknown temperature parameter,⁶ shown in Fig. 1(a). He then computes these two parameters by applying the laws of conservation of momentum and kinetic energy. It is to be expected that this procedure should give reasonable values for the mobility and the mean energy of the ion; indeed, if we consider the polarization force only, we get *exactly* the right values; the reason for this is that one may evaluate velocity averages for inverse fifth power forces ignoring the distribution function⁷ and that he did this in effect for the drift velocity and the

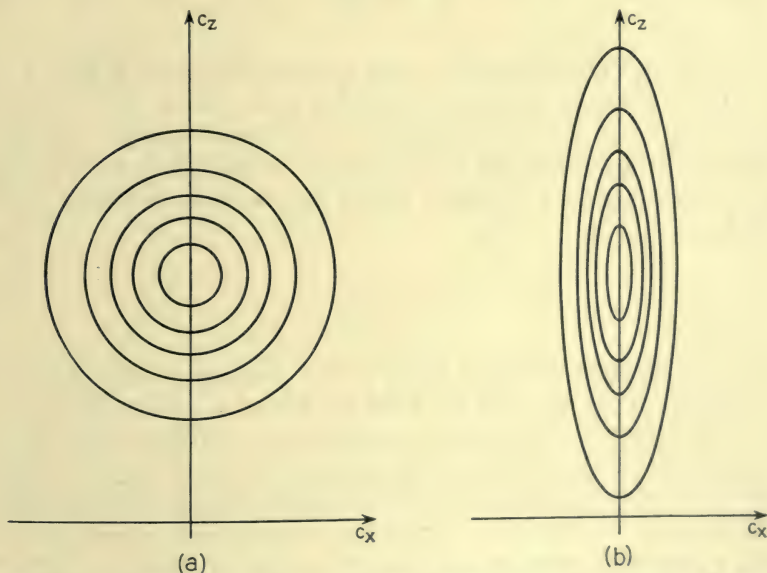


Fig. 1 — Simplified pictures for the high field velocity distribution of gaseous ions. (a) Hershey's assumption. (b) Modification with correct second moments.

⁶ Hershey, A. V., Phys. Rev., **56**, p. 916, 1939.

⁷ This will be shown in Section IIB.

total energy. In order to test whether an offset Maxwellian distribution is a satisfactory approximation we have to go one step further and examine the partition of the energy among the three degrees of freedom. There we find Hershey's distribution in error, for he assumes equipartition for the random motion, while, in reality, the random energy parallel to the field is much higher than at right angles,⁷ giving the distribution a decided "ridge" structure. This discrepancy could be taken into account by the use of an elliptically distorted Maxwellian distribution, shown in Fig. 1(b), and this may prove to be convenient in some applications.

For a detailed knowledge of the distribution function it is necessary to specify the interaction between an ion and a molecule. This interaction can be, broadly speaking, summarized under three headings: (a) the polarization force, (b) the short distance repulsion, and (c) symmetry effects. The polarization force arises because an ion, when passing close to a molecule, induces on it a dipole moment; this moment is then attracted by the charge of the ion. The attractive force F resulting from this is

$$F = \frac{2e^2P}{\rho^5} \quad (4)$$

where P is the polarizability of a gas molecule and e the charge of the ion. The force varies inversely as the fifth power of the distance ρ ; for such a force the cross section σ varies inversely as the speed of encounter γ . Whenever the cross section shows this type of variation it is advantageous to define a mean free time τ rather than a mean free path λ . The formula is

$$\tau = \frac{1}{N\sigma\gamma} \quad (5)$$

There is a standard difficulty which arises when one tries to make use of a formula of the type (5). For most force laws, a total cross-section σ cannot be defined; a differential cross section per unit solid angle always exists, but it becomes infinite in the forward direction because of small deflections suffered by particles passing by each other at a large distance. Thus equation (5) is, strictly speaking, meaningless. This is actually never a difficulty in the computation of a physical quantity. However, equation (5) is convenient for order-of-magnitude thinking and the question arises how it can be reasonably interpreted. The general method of salvaging (5) — excluding a small forward cone from consideration — is of little value for this purpose. An analysis of the inverse fourth power

attractive potential shows a better way out. The potential gives rise to two kinds of orbits; orbits of large angular momentum which look somewhat like hyperbolas, shown in Fig. 2(a), and orbits of small angular momentum for which the particles are "sucked" toward each other in a spiralling movement until a repulsive force reverses the trend, as shown in Fig. 2(b).⁸ A calculation of Hassé⁹ shows that the latter type of motion is much more efficient in scattering than the former and one gets therefore a picture which is semiquantitatively correct if one substitutes into (5) the cross-section for spiralling collisions and assumes isotropic scattering.¹⁰ This cross section equals

$$\sigma = 2\pi \sqrt{\frac{1}{m} + \frac{1}{M}} \frac{\sqrt{P} e}{\gamma} \quad (6)$$

A numerical estimate of the cross section (6) automatically leads one to compare it with the short distance repulsion familiar from the kinetic theory of gases. The two are of the same order, but for the usual gaseous speeds (which enter into (6) through γ) and small molecules the cross

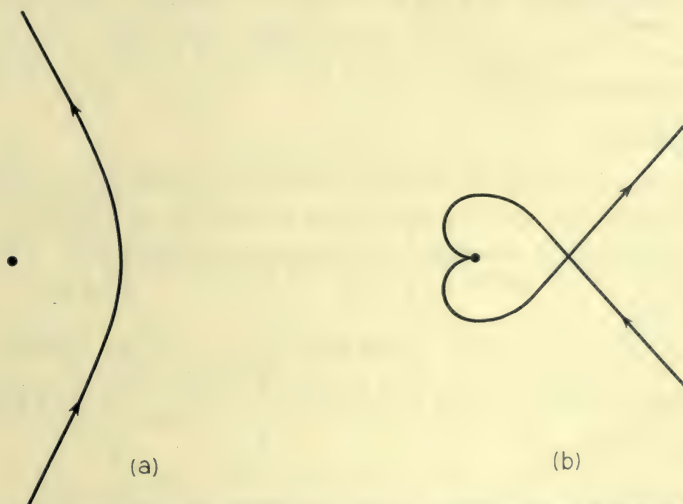


Fig. 2 — Sample orbits (schematic) showing the motion of a particle in the polarization force field. (a) Hyperbolic orbit (large angular momentum). (b) Spiralling orbit (small angular momentum).

⁸ There are quantum mechanical analogues to these classical ideas; they should lead to practically identical answers unless the angular momentum quantum number is small.

⁹ Hassé, H. R., *Phil. Mag.*, **1**, p. 139, 1926.

¹⁰ This will be discussed more fully in Section IIIB.

section (6) is bigger. This situation is accentuated in an actual scattering calculation which shows an attractive force to be generally more efficient than a repulsive force of equal range.

A detailed numerical discussion of these questions is found in Massey and Mohr¹¹ for the case of He^+ ions moving through He gas. Their interest is in the low field mobility. They show that for this problem the repulsive force makes so little difference that it could be neglected entirely without much affecting the results. It does finally come out that the polarization force gives a mobility which is too big by a factor of two. But the additional scattering is due to an effect which we listed above under (c): namely a resonance attraction between the He atom and the He^+ ion for which the cross section is abnormally large. It should be possible to eliminate this effect by increasing the cross section (6) until it masks even this special effect. Lowering the field is not sufficient to achieve this because of the temperature motion; it would be necessary in addition to reduce the absolute temperature by a sizeable factor and so to decrease the value of γ in (6). Thus we are led to the prediction that if the temperature of He is reduced the mobility of He^+ ions in He should gradually rise from its "anomalous" value of $12 \text{ cm}^2/\text{volt sec}$ to the "normal" value of $22 \text{ cm}^2/\text{volt sec}$, which one gets by taking account of polarization forces only.

IB. GLOSSARY

The complicated appearance of equations in gaseous kinetics suggests special care in the use of symbols and a convenient arrangement for the reader to find their meaning. It is hoped that the glossary to follow will accomplish this purpose. It explains all symbols except those used at one location only.

Generally, Latin capital letters will refer to the gas molecules and Latin lower case letters to the ions; Greek letters will have no special relationship; exceptions will be made for generally recognized symbols. Thus we define

$E, \mathbf{E}$ = electric field.

x, y = cartesian coordinates at right angles to the field direction.

z = cartesian coordinate along the field direction.

$\mathbf{r}$ = position vector with components x, y, z .

t = time.

m = ionic mass.

e = ionic charge.

¹¹ Massey, H. S. W., C. B. O. Mohr, Proc. Roy. Soc., **144A**, p. 554, 1931.

$a, \mathbf{a} = \frac{e\mathbf{E}}{m}$ = ionic acceleration.

b = impact parameter.

b_{lim} = limiting value of the impact parameter separating hyperbolic and spiralling orbits (equation (125)).

$c, \mathbf{c}, \mathbf{c}', c_x, c_f, c_i$ = various ionic velocities or components.

e_x, e_y, e_z = energies of ionic motion (or "high field" parts thereof) along x, y, z .

e_z^* = random part of the above energy,

$\mathbf{j}_i$ = total particle current density of the ions (may be a function of $\mathbf{r}$ and t).

$\mathbf{j}$ = partial current density induced by the concentration gradient; see equation (22).

k = Boltzmann's constant (only when followed directly by T).

$k, \mathbf{k}$ = relative concentration gradient of the ions (a different use is made of k in Section IIIE).

n = number density of the ions (may be a function of $\mathbf{r}$ and t).

p, q, r, s = undetermined constants; used three times independently (equations (82), (90) and (160)).

$p^{(0)}, p^{(1)}$ = various approximations to these numbers.

$\mathbf{u}, \mathbf{u}', \mathbf{v}$ = ionic velocities.

$\mathbf{w}$ = ionic velocity rendered dimensionless (see eq. (75) or (85)).

$\mathcal{J}$ = the inner integral in the double integral eq. (69).

$\mathbf{C}, \mathbf{C}'$ = molecular velocities.

$\mathcal{D}$ = ionic diffusion tensor.

$D_{||}, D_{\perp}$ = components of above tensor parallel and perpendicular to the field.

M = molecular mass.

N = number density of the molecules.

P = molecular polarizability.

T = gas temperature.

$\mathbf{U}, \mathbf{U}'$ = molecular velocities.

X, Y = left and right hand sides of equation (111a).

$\beta = \frac{1}{2kT}$ = temperature parameter (A different use of β is made

in Section IIIB where it is the relative impact parameter b/b_{lim}).

$\gamma, \gamma', \eta, \eta'$ = relative velocities of ion and molecule.

ξ, η, ζ = cartesian coordinates oriented on $\mathbf{c}$.

ρ = distance between ion and molecule.

σ = collision cross section of ions and molecules (may be a function of γ).

$\lambda = \frac{1}{N\sigma}$ = mean free path of ion between collisions with molecules
(may be a function of γ).

$\tau = \frac{1}{N\sigma\gamma}$ = mean free time for the ion between collisions with molecules.

τ_s = same parameter for "spiralling" collisions.

$\alpha = \frac{d \ln \tau(\gamma)}{d \ln \gamma} + 1$. It is assumed constant in Section ID.

χ, χ_c, χ_u = angle of scattering of ion and molecule in the center of mass system.

κ = angle of scattering of the ion by a molecule in the laboratory system.

ϵ = scattering azimuth of ion and molecule in the center of mass system.

ω = scattering azimuth in the laboratory system (azimuth of the initial ion velocity about the final ion velocity).

ϑ, ϑ' = angle between velocity vector and field direction.

$\psi, \varphi, \theta, \phi, \delta$ = other angles (these angles are defined on spherical triangles which are exhibited in Figs. 8 and 15).

$d(\mathbf{c}, \mathbf{r}, t)$ = density function of ions in phase space.

$m(\mathbf{c}) = \left(\frac{\beta m}{\pi}\right)^{3/2} \exp(-\beta m c^2)$ = Maxwellian velocity distribution function for ionic mass.

$M(\mathbf{C}) = \left(\frac{\beta M}{\pi}\right)^{3/2} \exp(-\beta M C^2)$ = Maxwellian velocity distribution function for molecular mass.

$h(\mathbf{c})$ = "high field" distribution function of the ions for the case that the spatial distribution is uniform (the exact meaning of this term is to be explained in the text).

$f(\mathbf{c})$ = true velocity distribution of the ions for the case that the spatial distribution is uniform.

$g(\mathbf{c})$ = correction to $f(\mathbf{c})$ or $h(\mathbf{c})$ for the case of a constant relative concentration gradient $\mathbf{k}$.

$\delta(\mathbf{c})$ = vectorial δ -function in velocity space.

$Ei(x) = \int_x^\infty \frac{e^{-\xi}}{\xi} d\xi$ (suppression of two minus signs).

$I_0(x)$ = modified Bessel function of order 0.

$K_0(x), K_1(x)$ = Modified Hankel functions of order 0, 1. (Alteration of Macdonald function by a factor $\frac{2}{\pi}$).

$P_n(x)$ = Legendre Polynomials.

$h_r(c), g_r(c)$ = expansion coefficients which result when $h(c), g(c)$ are expanded in Legendre Polynomials about the field direction.

$I_{s,r}(\chi)$ = A set of functions of the scattering angle defined in (48).

$\langle \quad \rangle$ = the quantity in pointed brackets is to be averaged.

$\langle s, \nu \rangle$ = abbreviation for $\langle w^s P_\nu(\cos \vartheta) \rangle$; the average is taken over $h(w)$.

$\{s, \nu\}$ = A normalized correction to $\langle s, \nu \rangle$ contributed by $g(w)$; see equation (155).

A special convention will be adopted to distinguish velocities before and after a collision:

c', C' = velocities before the collision.

c, C = velocities after the collision.

When used in this fashion the twelve components of the four vectors above satisfy the four identities:

$$mc' + MC' = mc + MC \quad (7)$$

$$mc'^2 + MC'^2 = mc^2 + MC^2 \quad (8)$$

The same convention is to apply to other vector quadruples, such as

$$u, U, u', U'$$

For the velocities in the center of mass system we use

$$\gamma' = c' - C' = \text{relative velocity before the collision.}$$

$$\gamma = c - C = \text{relative velocity after the collision.}$$

In consequence of (7) and (8) the γ 's obey the relation

$$\gamma'^2 = \gamma^2 \quad (9)$$

The multiple integrations occurring in the theory are of the following two types. Either they are over the three components of a velocity in a Cartesian velocity space; we shall denote such integrations by $d\mathbf{c}, d\mathbf{u}, d\mathbf{U}'$, etc. Or they are proper "collision" integrations which classically have the form

$$\gamma b db d\epsilon$$

where b is an impact parameter and ϵ an azimuth. In most cases these integrals depend on extraneous factors for their convergence but this fact is usually disregarded for convenience; we shall follow this habit by

writing the above differential in the form

$$\frac{1}{4\pi} \gamma \sigma(\gamma) \Pi(\chi) \sin \chi \, d\chi \, d\epsilon = \frac{1}{4\pi} \gamma \sigma(\gamma) \Pi(\chi) \, d\Omega_\gamma$$

Here $d\Omega$ is meant to represent an integration over a solid angle and the subscript γ , that it is over the solid angle swept out by the vector γ . The notation makes use of the fact that the choice of the polar axis is arbitrary in such an integration. The function $\Pi(\chi)$ is the probability of scattering which equals unity for isotropic scattering. In cases where small angle scattering is infinitely probable the above expression becomes meaningless, strictly speaking, $\Pi(\chi)$ being a δ -function at $\chi = 0$ and σ being infinite. However if a quantity such as $1 - \cos \chi$ is multiplied in, which removes the δ -function then the integration gives a finite number which may be denoted by $\langle \sigma \cdot (1 - \cos \chi) \rangle$.

IC. FORMAL SURVEY OF THE THEORY

Under the assumptions stated in Part IA we may describe the motion of ions in a gas by their density in phase space. The change in time of this function is described by a Boltzmann equation¹² which, in our notation, reads

$$\begin{aligned} \frac{\partial d(\mathbf{c}, \mathbf{r}, t)}{\partial t} + \mathbf{a} \cdot \frac{\partial d(\mathbf{c}, \mathbf{r}, t)}{\partial \mathbf{c}} + \mathbf{c} \cdot \frac{\partial d(\mathbf{c}, \mathbf{r}, t)}{\partial \mathbf{r}} \\ = \frac{N}{4\pi} \iint \{M(\mathbf{C}')d(\mathbf{c}, \mathbf{r}, t) - M(\mathbf{C})d(\mathbf{c}, \mathbf{r}, t)\} \gamma \sigma(\gamma) \Pi(\chi) \, d\Omega_\gamma \, d\mathbf{C} \end{aligned} \quad (10)$$

The equation is linear in the unknown function $d(\mathbf{c}, \mathbf{r}, t)$; this is due to neglect of ion-ion collisions, as stated earlier. The negative term on the right hand side actually reduces to a known function of $\mathbf{c}$ multiplying $d(\mathbf{c}, \mathbf{r}, t)$. The positive term is a genuine integral term; it has been shown by Pidduck¹³ that the number of integrations in it can be brought down from five to three; this reduction will not be made use of in the following.

If there were no terms on the left hand side of equation (10) then the solution of it would have the equilibrium form

$$d(\mathbf{c}, \mathbf{r}, t) = nm(\mathbf{c}) \quad (11)$$

where n is a constant. This result is a direct consequence of equation (8) which makes the curly bracket in (10) vanish identically when Maxwellian functions are inserted.

¹² See Reference 4.

¹³ Pidduck, F. B., Proc. Lond. Math. Soc., **15**, p. 89, 1915.

The function $m(\mathbf{c})$ is not the solution of our problem because of the presence of the second and third term on the left which arise from an electric field and a density variation respectively. These disturbances will be assumed of different relative importance. The density variation will be assumed sufficiently small so that the third term can be treated by perturbation theory; the field term, on the other hand will be taken so large that the equilibrium distribution (11) no longer represents a first approximation to the solution. In consequence, the equation is solved in two stages. In the first, only the second term on the left is retained, and the resultant equation is treated rigorously; in the second, the full equation (10) is used, but the new terms are taken as perturbations.

The first stage describes those properties of the ion gas which it possesses when assumed of uniform density. Since the field is also assumed uniform and not changing in time, the dependence on $\mathbf{r}$ and t drops out. We may then write

$$d(\mathbf{c}, \mathbf{r}, t) = n f(\mathbf{c}) \quad (12)$$

where n is a constant and $f(\mathbf{c})$ is a velocity distribution function. The equation for f reads

$$\mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{c}} = \frac{N}{4\pi} \iint \{M(\mathbf{C}')f(\mathbf{c}') - M(\mathbf{C})f(\mathbf{c})\} \gamma \sigma(\gamma) \Pi(\chi) d\Omega_{\gamma'} d\mathbf{C} \quad (13)$$

with the side condition

$$\int f(\mathbf{c}) d\mathbf{c} = 1 \quad (14)$$

As a result of solving (13) we shall obtain the distribution function $f(\mathbf{c})$ as a function of the electric field contained in $\mathbf{a}$. This distribution differs essentially from the Maxwellian one in that it is not symmetric about the origin. The vectorial mean of the velocity is therefore not zero

$$\langle \mathbf{c} \rangle = \int f(\mathbf{c}) \mathbf{c} d\mathbf{c} \neq 0 \quad (15)$$

This is the drift velocity of the ion in the field which is reached as a compromise between the acceleration $\mathbf{a}$ and the frictional losses caused by the ion-atom collisions. From the structure of equation (13) there is one general prediction that can be made concerning this velocity, namely that it depends on the gas density and the field only through $\mathbf{a}/N$; this is the well known E/p_0 of the experimental analysis. This type of

dependence does not only hold for $\langle \mathbf{c} \rangle$, but for all averages derivable from $f(\mathbf{c})$, notably the mean energy.

A more important formal prediction can be made about the second stage of the contemplated calculation. For it will be shown now that the diffusion concept is still applicable in the presence of a strong electric field. It is true, that if we have a variable density in space the primary motion observed is not diffusive but a displacement of the entire density pattern with the drift velocity $\langle \mathbf{c} \rangle$. However, once this dominant component is subtracted out, then a supplementary current proportional to the density gradient is identified. The constant of proportionality is anisotropic, that is, we have a diffusion tensor rather than a diffusion coefficient. The tensor is axially symmetric about the field direction, yielding a longitudinal and a transverse diffusion coefficient.

To demonstrate these features it is convenient to assume a special type of variation of ion density in space. As we shall see the velocity distribution is primarily sensitive to the relative density gradient $\mathbf{k}$; we shall therefore assume it to be a constant. In other words we set

$$n(\mathbf{r}, t) = n_0 \exp [\mathbf{k} \cdot (\mathbf{r} - \langle \mathbf{c} \rangle t)] \quad (16)$$

The relation can of course not hold everywhere since n increases beyond all bounds in one direction, but we must remember here that we are doing perturbation theory, that is $\mathbf{k}$ is assumed small. The inconsistencies in the assumption (16) can then be pushed as far away as we please. Furthermore there is no inconsistency at all in the half space where n decreases. It is to be observed that according to the assumption (16) the spatial distribution is moving unchanged through space with the drift velocity $\langle \mathbf{c} \rangle$. This seems to contradict the program of finding the effect of diffusion upon $n(\mathbf{r}, t)$. However, we follow in this simply conventional steady state computational methods in which a gradient is assumed maintained from an infinitely strong source; the modification appears then as a change in the velocity distribution function, which, in turn, yields a steady diffusion current. We set therefore

$$d(\mathbf{c}, \mathbf{r}, t) = n(\mathbf{r}, t)[f(\mathbf{c}) + g(\mathbf{c})] \quad (17)$$

where $g(\mathbf{c})$ is a correction to the solution of (13) which arises from the assumption (16). It follows from the definition of $n(\mathbf{r}, t)$ and (14) that

$$\int g(\mathbf{c}) d\mathbf{c} = 0 \quad (18)$$

The consistency of the assumptions (16) and (17) with equation (10) becomes evident when they are substituted into this equation. We

find, after simplification with (13)

$$\begin{aligned} \mathbf{a} \cdot \frac{\partial g(\mathbf{c})}{\partial \mathbf{c}} + \frac{N}{4\pi} \iint \{M(\mathbf{C})g(\mathbf{c}) - M(\mathbf{C}')g(\mathbf{c}')\} \gamma \sigma(\gamma) \Pi(\chi) d\Omega_{\gamma'} d\mathbf{C} \\ = -\mathbf{k} \cdot (\mathbf{c} - \langle \mathbf{c} \rangle) \{f(\mathbf{c}) + g(\mathbf{c})\} \end{aligned} \quad (19)$$

This is an equation in velocity space only, $\mathbf{r}$ and t having disappeared completely; this justifies the assumptions. In solving the equation we observe that our interest is only in diffusion, that is, the current resulting from a concentration gradient when treated in first order perturbation. In this case both $\mathbf{k}$ and $g(\mathbf{c})$ are to be treated as small and their product in (19) is to be neglected. The equation then becomes

$$\begin{aligned} \mathbf{a} \cdot \frac{\partial g(\mathbf{c})}{\partial \mathbf{c}} + \frac{N}{4\pi} \iint \{M(\mathbf{C})g(\mathbf{c}) - M(\mathbf{C}')g(\mathbf{c}')\} \gamma \sigma(\gamma) \Pi(\chi) d\Omega_{\gamma'} d\mathbf{C} \\ = -\mathbf{k} \cdot (\mathbf{c} - \langle \mathbf{c} \rangle) f(\mathbf{c}) \end{aligned} \quad (20)$$

The homogeneous prototype of this inhomogeneous equation is (13); an arbitrary amount of $f(\mathbf{c})$ could thus be added to a particular solution of (20) were it not for the orthogonality condition (18) which makes the solution definite.

The existence of the diffusion phenomenon follows easily from equation (20). The total current $\mathbf{j}_t$ is given by

$$\mathbf{j}_t(\mathbf{r}, t) = \int d(\mathbf{c}, \mathbf{r}, t) \mathbf{c} d\mathbf{c} \quad (21)$$

Upon substitution of (17) into this expression two terms result

$$\mathbf{j}_t(\mathbf{r}, t) = n(\mathbf{r}, t) \langle \mathbf{c} \rangle + \mathbf{j}(\mathbf{r}, t) \quad (22)$$

with

$$\mathbf{j}(\mathbf{r}, t) = n(\mathbf{r}, t) \int g(\mathbf{c}) \mathbf{c} d\mathbf{c} \quad (23)$$

The first term in (22) is seen from (15) to just equal the product of the density and the drift velocity; this is the expected drift current. The new current $\mathbf{j}(\mathbf{r}, t)$ induced by the density gradient is thus given by (23). From (20) it follows that $g(\mathbf{c})$ is a linear function of the three components k_x, k_y, k_z with coefficients which do not depend on the density or its gradient, but only on the unperturbed velocity distribution $f(\mathbf{c})$; furthermore, the first two of these coefficients are equal. Hence, from (23) $\mathbf{j}$ comes out as a linear function of the three quantities $n(\mathbf{r}, t) \cdot k_x, n(\mathbf{r}, t) \cdot k_y, n(\mathbf{r}, t) \cdot k_z$; these are the components of the density gradient

as is evident from (16); in addition the multipliers of the first two components are equal. We may write therefore

$$\mathbf{j}(\mathbf{r}, t) = -(\mathcal{D}) \frac{\partial n}{\partial \mathbf{r}} \quad (24)$$

where $(\mathcal{D})$ is a tensor which is axially symmetric about the field direction; its two components which we shall call the longitudinal diffusion coefficient $D_{||}$ and the transverse coefficient $D_{\perp}$ are computed entirely from the unperturbed velocity distribution $f(\mathbf{c})$. This makes $D_{||}$ and $D_{\perp}$ independent of the density or its gradient; it is to be noted, however, that they do depend on the electric field as a parameter because this quantity enters several times in the course of the computation.

ID. DIMENSIONAL ANALYSIS

Dimensional analysis is a convenient tool in a qualitative discussion of (13) and (20). In order to get results the situation has to be schematized somewhat, but not so much as to impair its usefulness. In the first place it is convenient to keep in mind the two limiting cases of high and low field, as discussed in the introduction. In addition some assumption must be made about $\sigma(\gamma)$ and $\Pi(\chi)$ occurring under the integral sign. The most convenient way to dispose of $\Pi(\chi)$ is to take it as independent of γ . This happens to be true for the two models treated in detail later, the polarization force model, and the hard sphere model. Actually $\Pi(\chi)$ can be taken as approximately independent of γ in a wider sense. The forces which produce scattering are either repulsive or short range attractive, that is, long range attractive forces are absent. As long as this is the case the scattering is roughly isotropic and hence can change but little with γ .¹⁴

A more drastic assumption is needed to dispose of $\sigma(\gamma)$. We must assume

$$\sigma(\gamma) \cdot \gamma^{\alpha} = \Gamma \quad (25)$$

where α and Γ are taken to be constants. This assumption contains two important special cases in it. They arise respectively by taking $\alpha = 0$ and $\alpha = 1$. The case $\alpha = 0$ is the case of a constant mean free path as exemplified by the hard sphere model. The case $\alpha = 1$ is the case of constant mean free time; it is applicable to the polarization force as dis-

¹⁴ This statement is checked in detail in Section IIIB for the polarization force. This is the attractive force with the longest range which can arise in this field.

cussed in Section IA. When (25) is inserted into (13) it is seen that a , N and Γ enter only in the combination $a/N\Gamma$. The quantity

$$\left(\frac{a}{N\Gamma}\right)^{\frac{1}{2-\alpha}}$$

has the dimension of a velocity. A second such quantity is

$$\left(\frac{kT}{M}\right)^{1/2}$$

which arises from the Maxwellian functions under the integral sign.¹⁵ In the high field case, this quantity does not enter, that is, the velocity distribution functions for the molecules could be replaced by δ -functions at the origin. Hence the first combination controls all velocity averages. For the mean drift velocity, we can thus write

$$\langle c_z \rangle = \text{const} \cdot \left(\frac{a}{N\Gamma}\right)^{\frac{1}{2-\alpha}} \quad (26a)$$

This formula gives the variation of the drift velocity with the electric field. It is worth while writing the result out explicitly for the two special cases discussed above. The first is the case of constant mean free path, $\alpha = 0$, for which

$$\langle c_z \rangle = \text{const} \cdot a^{1/2} \lambda^{1/2} \quad (26b)$$

This is a drift velocity varying as the square root of the field or a mobility varying inversely as the square root of the field. The second case is the one of constant mean free time $\alpha = 1$, for which

$$\langle c_z \rangle = \text{const} \cdot a\tau \quad (26c)$$

This means a drift velocity proportional to the field or a constant mobility.

In the low field case we cannot disregard one of the two velocity parameters constructed above; but now equation (13) is to be solved by perturbation theory only, it then yields a drift velocity varying with the first power of $a/N\Gamma$. Dimensional analysis then yields the dependence of the mobility on the temperature. We find

$$\langle c_z \rangle = \text{const} \cdot \frac{a}{N\Gamma} \left(\frac{kT}{M}\right)^{\frac{\alpha-1}{2}} \quad (27a)$$

¹⁵ Dimensional analysis is incapable of distinguishing between m and M ; this means that we cannot master dependence on mass by the method of this section; all our "pure numbers" are actually unknown functions of m/M .

with the special cases

$$\langle c_z \rangle = \text{const} \cdot a \lambda \left(\frac{kT}{M} \right)^{-1/2} \quad (27b)$$

for constant mean free path and

$$\langle c_z \rangle = \text{const} \cdot a \tau \quad (27c)$$

for constant mean free time. Comparison of (26c) and (27c) might lead one to surmise that we have here twice the same formula. This is indeed the case, as will be shown in Section IIIA.

Proceeding now to the diffusion problem, we observe from (20), (23) and (24) that we must add the quantity

$$\frac{nk}{N\Gamma}$$

to the previous list of parameters when computing the diffusion current. However, the current is always linear in this quantity, which means that the diffusion coefficients contain the factor

$$\frac{1}{N\Gamma}$$

and beyond this factor depend on the same variables as previously. This gives in the high field case

$$D = \text{const} \cdot \frac{1}{N\Gamma} \cdot \left(\frac{a}{N\Gamma} \right)^{\frac{1+\alpha}{2-\alpha}} \quad (28a)$$

with the special cases

$$D = \text{const} \cdot a^{1/2} \lambda^{3/2} \quad (28b)$$

and

$$D = \text{const} \cdot a^2 \tau^3 \quad (28c)$$

In the low field case, the diffusion process becomes independent of the field and we get

$$D = \text{const} \cdot \frac{1}{N\Gamma} \left(\frac{kT}{M} \right)^{\frac{1+\alpha}{2}} \quad (29a)$$

with the special formulas

$$D = \text{const} \cdot \lambda \left(\frac{kT}{M} \right)^{1/2} \quad (29b)$$

and

$$D = \text{const} \cdot \tau \frac{kT}{M} \quad (29c)$$

The information in the formulas (29) is now new, but dependent on (27) through a universal relation first discovered by Nernst and derived independently for gases by J. J. Thomson; it is widely known as the Einstein relation. It states that

$$D = \frac{\partial \langle c_z \rangle}{\partial a} \cdot \frac{kT}{m} \quad (30)$$

Equation (30) contains of course more than is obtainable from (27) and (29), since it relates one undetermined constant to another in a known way.

The dimensional methods of this section are convenient for a rough classification of experimental material. Figs. 3 to 7 show the drift velocities

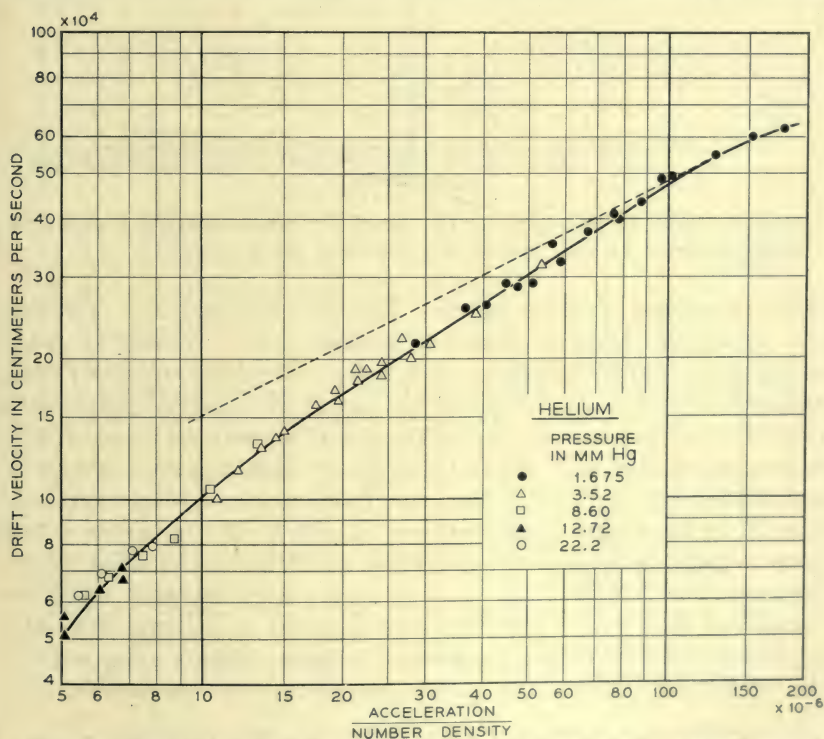


FIG. 3 — Drift velocity in an electric field of He^+ ions in helium gas. Comparison of observed results with an “asymptotic” straight line of slope $\frac{1}{2}$.

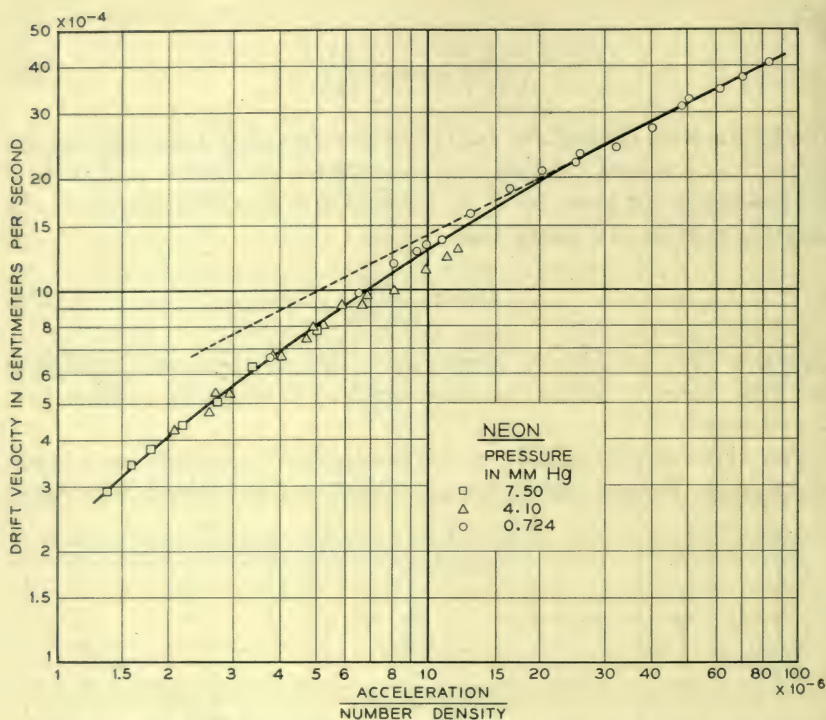


FIG. 4 — Drift velocity in an electric field of Ne^+ ions in neon gas. Comparison of observed results with an "asymptotic" straight line of slope $\frac{1}{2}$

in the parent gas, observed for He^+ , Ne^+ , A^+ , Kr^+ and Xe^+ . The plot is a log-log plot of these quantities against a/N , a variety of fields having been used to determine each point. The data are taken from measurements of J. A. Hornbeck^{16, 17} and R. N. Varney.¹⁸ These data verify in the first place that the drift velocity depends on a and N only in the combination a/N . Beyond this we see that the curves consist of two straight line portions: in the lower field portion $\langle c_z \rangle$ is proportional to a/N , in the higher is proportional to $\sqrt{a/N}$. We recognize in this latter region the high field dependence predicted in equation (26b). We learn from this that the collision cross section between noble gas atoms and their ions is approximately constant in the experimentally significant velocity range. To determine these collision cross sections the computation of only a single number, namely the one entering into

¹⁶ Hornbeck, J. A. and G. H. Wannier., Phys. Rev., **82**, p. 458, 1951.

¹⁷ Hornbeck, J. A., Phys. Rev., **84**, p. 615, 1951.

¹⁸ Varney, R. N., Phys. Rev., **88**, p. 362, 1952.

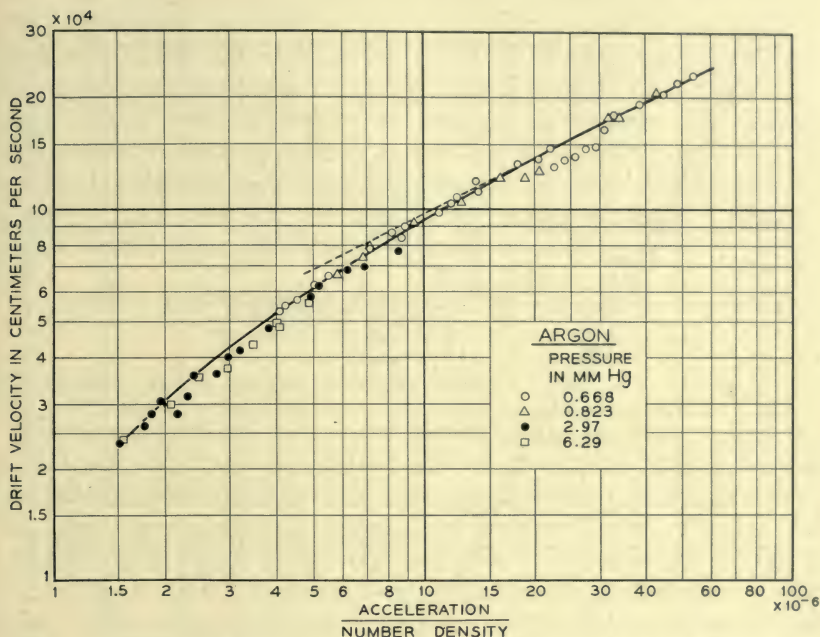


FIG. 5 — Drift velocity in an electric field of A^+ ions in argon gas. Comparison of observed results with an “asymptotic” straight line of slope $\frac{1}{2}$.

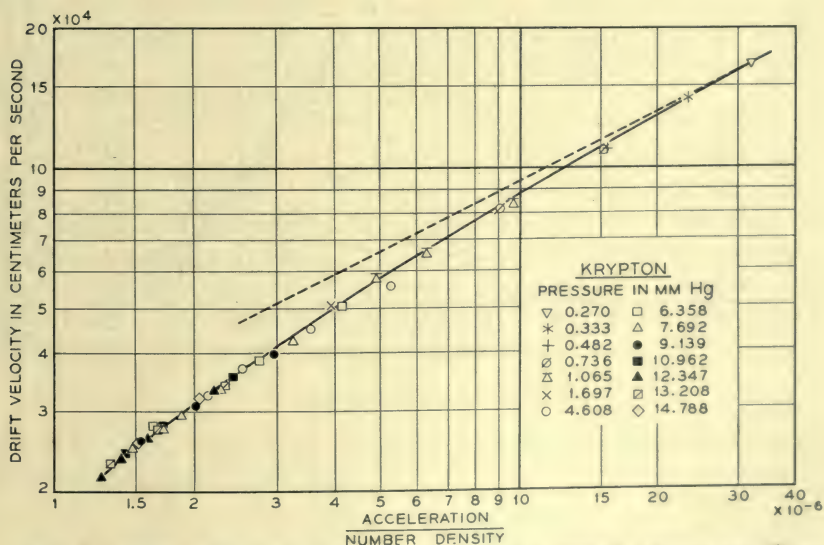


FIG. 6 — Drift velocity in an electric field of Kr^+ ions in krypton gas. Comparison of observed results with an “asymptotic” straight line of slope $\frac{1}{2}$.

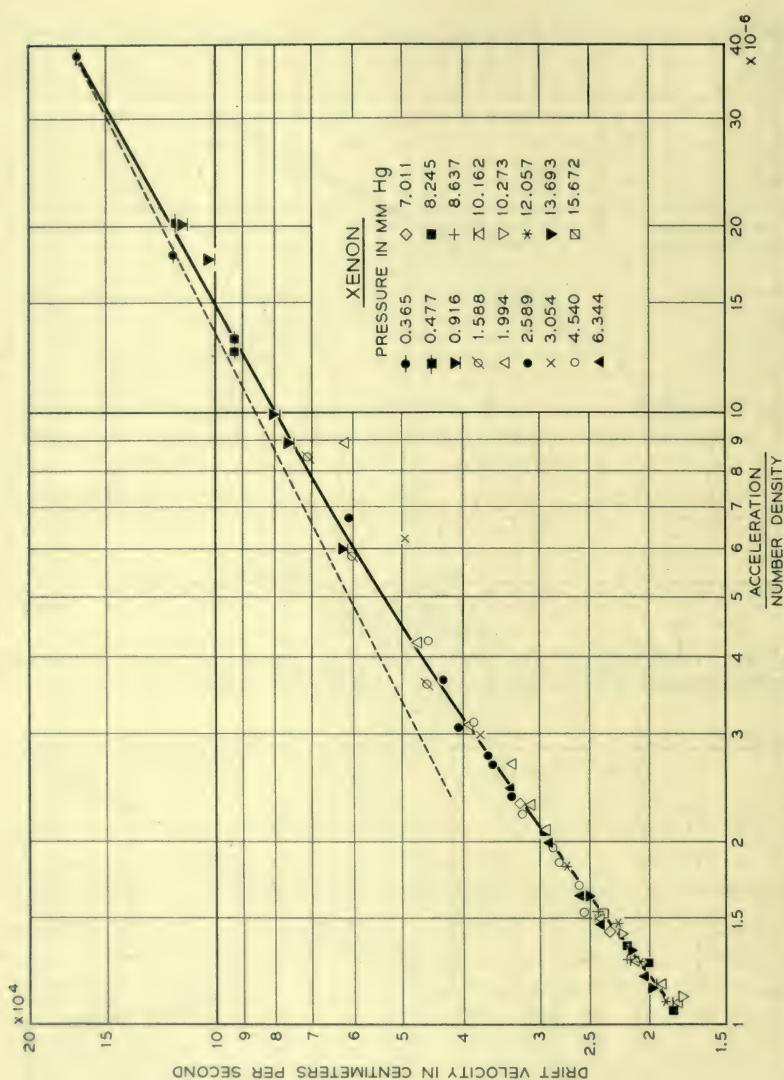


Fig. 7 — Drift velocity in an electric field of Xe^+ ions in xenon gas. Comparison of observed results with an "asymptotic" straight line of slope $\frac{1}{2}$.

(26b) is required. The linear range of this plot is not as informative as the high field one. The slope unity is common to all formulas (27), and the temperature dependence of the mobility is needed to give the correct interpretation with the methods developed here. There is a certain likelihood that the parameter α of equation (25) drifts from 0 to 1 as the speed of the ions is reduced; this was pointed out for the special case of He^+ in He in section IA. A qualitatively similar situation appears to prevail for the other noble gases.

PART II — THE MOTION OF UNIFORM ION STREAMS IN THE HIGH FIELD CASE

IIA. FORMULATIONS OF THE BOLTZMANN EQUATION

The dimensional analysis of the last section shows that there is an intrinsic simplicity to the high field case which is comparable to the low field case, while the intermediate case is more difficult. With one exception,⁶ however, theoretical analysis has occupied itself with the low field case only. We shall try to remedy this in the following. To begin with, a tractable but accurate formulation of the problem has to be found. Such a formulation cannot treat the field term of equation (13) as a perturbation term, but must try instead to make use of the basically simple features of the problem, notably those exhibited by the dimensional analysis of Section ID.

The equation governing the high field properties of the ions is obtained simply by substituting δ -functions for the Maxwellian velocity distributions in equation (13). This gives

$$a \frac{\partial h(\mathbf{c})}{\partial c_x} + \frac{1}{\tau(\mathbf{c})} h(\mathbf{c}) = \frac{1}{4\pi} \iint \delta(\mathbf{C}') h(\mathbf{c}') \frac{1}{\tau(\mathbf{c}')} \Pi(\chi) d\Omega_{\gamma'} d\mathbf{C} \quad (31)$$

A reduction of the number of integrations from five to two must be possible in the integral term of (31), owing to the presence of the δ -function. To achieve this we must transform the variables of integration so as to make three of the differentials equal to $d\mathbf{C}'$. We do this in the following way. First observe that

$$\boldsymbol{\gamma} = \mathbf{c} - \mathbf{C}$$

and that $\mathbf{c}$ is a constant vector. Hence we may replace $d\mathbf{C}$ by $d\boldsymbol{\gamma}$. The five-fold integration reads then

$$d\Omega_{\gamma'} d\mathbf{C} = \gamma^2 d\gamma d\Omega_{\gamma'} d\Omega_{\gamma'} \quad (32)$$

that is, it goes over the magnitude γ which the two vectors have in

common, and their orientations, for which they are independent. It is known that in integrating over the two angles defining an orientation the polar axis may be chosen freely. We shall, in the following, adopt $\mathbf{c}$ as our polar axis with ψ, κ being pole distances of γ and γ' to $\mathbf{c}$ and φ, ω the corresponding azimuths. In Fig. 8 these angles are exhibited on the unit sphere. All vectors are assumed to be plotted from the center of the sphere, and show up through their piercing points. The angles between the vectors then show up as sides and the azimuths as angles. The expression (32) becomes then

$$\gamma^2 d\gamma \sin \psi d\psi d\varphi \sin \kappa d\kappa d\omega$$

The main transformation consists now in introducing the three components of $\mathbf{C}'$ in the place κ, ψ and φ . The transformation formulas follow from the vector identity

$$\mathbf{C}' = \mathbf{c} - \frac{M}{M+m} \gamma - \frac{m}{M+m} \gamma' \quad (33)$$

and read in full

$$C'_x = -\frac{M\gamma}{M+m} \sin \psi \cos \varphi - \frac{m\gamma}{M+m} \sin \kappa \cos \omega$$

$$C'_y = -\frac{M\gamma}{M+m} \sin \psi \sin \varphi - \frac{m\gamma}{M+m} \sin \kappa \sin \omega$$

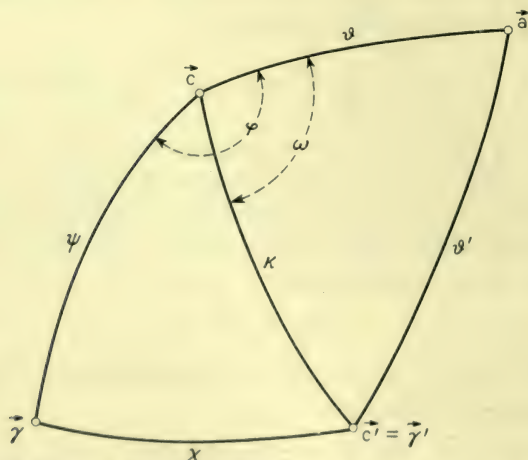


Fig. 8 — Definition of the angles employed in the formulations of the Boltzmann equation for the high field case.

$$C'_z = -\frac{M\gamma}{M+m} \cos \psi - \frac{m\gamma}{M+m} \cos \kappa + c$$

From these equations the value of the Jacobian comes out to be

$$\begin{aligned} & \frac{\partial(C'_\xi, C'_\eta, C'_z)}{\partial(\kappa, \psi, \varphi)} \\ &= \frac{mM^2\gamma^3}{(M+m)^3} \sin \psi \{ \cos \psi \sin \kappa - \sin \psi \cos \kappa \cos(\varphi - \omega) \} \end{aligned}$$

We need its value only at the position $C'_\xi = C'_\eta = C'_z = 0$. If we take the above equations for C'_ξ , C'_η , C'_z and multiply them respectively by $\cos \kappa \cos \omega$, $\cos \kappa \sin \omega$, $-\sin \kappa$, add and set $\mathbf{C}' = 0$ we get the identity

$$\frac{M\gamma}{M+m} \{ \cos \psi \sin \kappa - \sin \psi \cos \kappa \cos(\varphi - \omega) \} = c \sin \kappa$$

The curly bracket is exactly the one occurring in the Jacobian which therefore reduces to

$$\left[\frac{\partial(C'_\xi, C'_\eta, C'_z)}{\partial(\kappa, \psi, \varphi)} \right]_{\mathbf{C}'=0} = \frac{Mm}{(M+m)^2} c'^2 \sin \psi \sin \kappa$$

and hence

$$\gamma^2 d\gamma d\Omega_\gamma d\Omega_{\gamma'} = \frac{(M+m)^2}{Mmc} d\mathbf{C}' dc' d\omega$$

Substituting finally this expression into (32) and (31) we get the Boltzmann equation in the form

$$\begin{aligned} a \frac{\partial h(\mathbf{c})}{\partial c_z} + \frac{1}{\tau(\mathbf{c})} h(\mathbf{c}) &= \\ &= \frac{(M+m)^2}{4\pi Mmc} \int_c^{\frac{M+m}{|M-m|}c} \frac{1}{\tau(c')} \Pi(\chi) dc' \int_0^{2\pi} h(\mathbf{c}') d\omega \end{aligned} \quad (34)$$

The equation is in need of additional elucidation as regards the exact meaning of $\mathbf{c}'$ as a vector and as regards the auxiliary variable χ . As to the first point we may describe the integration as occurring over a surface in velocity space. This surface is obtained from the relation

$$\mathbf{C}' = 0 \quad \mathbf{c}' = \boldsymbol{\gamma}' \quad (35)$$

which substituted into (33) becomes

$$(M+m)\mathbf{c} - m\boldsymbol{\gamma}' = M\boldsymbol{\gamma} \quad (36)$$

Squaring this and using (9) we get

$$(M - m)c'^2 + 2mc' \cdot \mathbf{c} - (M + m)c^2 = 0 \quad (37)$$

This is the equation of a sphere in velocity space which passes through the point $\mathbf{c}' = \mathbf{c}$. For all other points c' is bigger than c (collision with a stationary object always brings energy loss). The center of the sphere lies on the line joining $\mathbf{c}$ to the origin; it lies on the side of the origin from $\mathbf{c}$ when $m < M$, at infinity (making the sphere a plane) when $m = M$ and away from the origin when $m > M$. We make use of (37) to express the polar angle κ of $\mathbf{c}'$ with respect to $\mathbf{c}$ (which does not occur as an integration variable in (34)) in terms of c' . We get

$$\cos \kappa = \frac{(M + m)c^2 - (M - m)c'^2}{2mcc'} \quad (38)$$

The angle of scattering in the center of mass system also results from squaring of (36) if the term $m\gamma'$ is first taken to the right. We find

$$\cos \chi = \frac{(M + m)^2}{2Mm} \frac{c^2}{c'^2} - \frac{M^2 + m^2}{2Mm} \quad (39)$$

There is a more useful form of equation (34) which results if χ is taken as one of the integration variables rather than c' . Substitution is made from the equation (39) above; it yields

$$a \frac{\partial h(\mathbf{c})}{\partial c_z} + \frac{1}{\tau(c)} h(\mathbf{c}) = \frac{1}{4\pi} \int_0^\pi \sin \chi d\chi \frac{\Pi(\chi)}{\tau(c')} \left(\frac{c'}{c}\right)^3 \int_0^{2\pi} h(\mathbf{c}') d\omega \quad (40)$$

The magnitude of $\mathbf{c}'$ and its polar angle with respect to $\mathbf{c}$ are now auxiliary parameters; the first is obtained from (39)

$$c' = c \frac{M + m}{\sqrt{M^2 + m^2 + 2Mm \cos \chi}} \quad (41)$$

and the second from (38) and (41)

$$\cos \kappa = \frac{m + M \cos \chi}{\sqrt{M^2 + m^2 + 2Mm \cos \chi}} \quad (42)$$

As previously, the azimuth ω of $\mathbf{c}'$ about $\mathbf{c}$ is an independent variable.

The simplifications of the equation (31) exhibited in (34) and (40) still leave a double integral in the fundamental equation. The integration over $d\omega$ will now be eliminated by decomposition of $h(\mathbf{c})$ in spherical harmonics about the field direction. There is no loss of generality in this step.

$$h(\mathbf{c}) = \sum_{\nu=0}^{\infty} h_\nu(c) P_\nu(\cos \vartheta) \quad (43)$$

We have now to consider simultaneously the three vectors $\mathbf{c}$, $\mathbf{c}'$ and $\mathbf{a}$ as well as the angles between them. These angles are defined in Fig. 8. We study equation (34) or (40) term by term in order to see what becomes of it upon substitution of (43). Starting with $h(\mathbf{c}')$ under the integral sign we get from Fig. 8 and the addition theorem for spherical harmonics

$$h(\mathbf{c}') = \sum_{\nu=0}^{\infty} h_{\nu}(\mathbf{c}') [P_{\nu}(\cos \vartheta) P_{\nu}(\cos \kappa) + 2 \sum_{\mu=1}^{\nu} \frac{(\nu - \mu)!}{(\nu + \mu)!} P_{\nu}^{\mu}(\cos \vartheta) P_{\nu}^{\mu}(\cos \kappa) \cos \mu \omega]$$

For this expression, the integration over ω is elementary and gives

$$\int_0^{2\pi} h(\mathbf{c}') d\omega = 2\pi \sum_{\nu=0}^{\infty} h_{\nu}(\mathbf{c}') P_{\nu}(\cos \vartheta) P_{\nu}(\cos \kappa) \quad (44)$$

Further, we get for the derivative in (34) or (40)

$$\begin{aligned} & \frac{\partial}{\partial c_z} \left(\sum_{\nu=0}^{\infty} h_{\nu}(c) P_{\nu}(\cos \vartheta) \right) \\ &= \sum_{\nu=0}^{\infty} \frac{dh_{\nu}(c)}{dc} \frac{1}{2\nu + 1} \{ (\nu + 1) P_{\nu+1}(\cos \vartheta) + \nu P_{\nu-1}(\cos \vartheta) \} \\ & \quad + \frac{1}{c} h_{\nu}(c) \frac{\nu(\nu + 1)}{2\nu + 1} \{ P_{\nu-1}(\cos \vartheta) - P_{\nu+1}(\cos \vartheta) \} \end{aligned} \quad (45)$$

Through the equations (43), (44) and (45), all terms in equation (34) or (40) are developed in spherical harmonics with respect to the angle ϑ between $\mathbf{c}$ and the field direction. We can therefore annul separately the coefficient of each Legendre polynomial in $\cos \vartheta$. This gives the following set of equations

$$\begin{aligned} & \frac{(M + m)^2}{2Mmc} \int_c^{\frac{M+m}{|M-m|}} e^{\frac{h_{\nu}(\mathbf{c}')}{\tau(\mathbf{c}')}} P_{\nu}(\cos \kappa) \Pi(\chi) d\mathbf{c}' - \frac{h_{\nu}(c)}{\tau(c)} \\ &= \frac{\nu a}{2\nu - 1} \left(\frac{dh_{\nu-1}(c)}{dc} - \frac{\nu - 1}{c} h_{\nu-1}(c) \right) \\ & \quad + \frac{(\nu + 1)a}{2\nu + 3} \left(\frac{dh_{\nu+1}(c)}{dc} + \frac{\nu + 2}{c} h_{\nu+1}(c) \right) \end{aligned} \quad (46)$$

or

$$\frac{1}{2} \int_0^\pi \frac{h_\nu(c')}{\tau(c')} \left(\frac{c'}{c}\right)^3 P_\nu(\cos \kappa) \Pi(\chi) \sin \chi d\chi$$

$$- \frac{h_\nu(c)}{\tau(c)} = \frac{\nu a}{2\nu - 1} \left\{ \frac{dh_{\nu-1}(c)}{dc} - \frac{\nu - 1}{c} h_{\nu-1}(c) \right\} \quad (47)$$

$$+ \frac{(\nu + 1)a}{2\nu + 3} \left\{ \frac{dh_{\nu+1}(c)}{dc} + \frac{\nu + 2}{c} h_{\nu+1}(c) \right\}$$

where

$$\nu = 0, 1, 2, 3 \dots$$

The auxiliary parameters entering are given by (38) and (39) for equation (46), and (41) and (42) for equation (47).

The equations (46) or (47) obtained by Legendre decomposition still are, in general, mixed integral-differential equations in one independent variable. Further simplification is possible only in special cases some of which will be discussed later. An even more simple and tractable form of the Boltzmann equation can be achieved in general, however, if one gives up the idea of determining the velocity distribution function and concentrates instead on its moments. In other words, the Boltzmann equation can be looked upon as a system of relations between velocity averages, and as such it becomes a linear algebraic system.

To carry out this reduction we multiply equation (47) by c^{s+2} and integrate from 0 to ∞ . The second term on the left is then a simple velocity average. The same is true on the right hand side if two integrations by part are permissible and leave no integrated out part. $s \geq -1$ is probably adequate for this. The integral over the integral term at first looks as follows

$$\frac{1}{2} \int_0^\infty c^{s+2} dc \int_0^\pi \frac{h_\nu(c')}{\tau(c')} \left(\frac{c'}{c}\right)^3 P_\nu(\cos \kappa) \Pi(\chi) \sin \chi d\chi$$

In this expression we pass from c to c' as the independent variable. From (41) we see that

$$\frac{dc}{c} = \frac{dc'}{c'}$$

Hence the expression becomes

$$\int_0^\infty c'^{s+2} \frac{h_\nu(c')}{\tau(c')} dc' \frac{1}{2} \int_0^\pi \left(\frac{c}{c'}\right)^3 P_\nu(\cos \kappa) \Pi(\chi) \sin \chi d\chi$$

From (41) and (42) it is seen that this is actually the product of two independent integrals if the angular distribution $\Pi(\chi)$ is independent of the velocity of encounter c' . The first integral is then identical with the one arising from the second term in (47), and the second is a collision integral having no connection with the velocity distribution. Even if this is not the case, the second integral is still a dynamic average which can be evaluated as a function of c' previously to any knowledge of $h(c')$. We express this by introducing the abbreviation

$$I_{s,\nu} = \left(\frac{c}{c'}\right)^s P_\nu(\cos \kappa)$$

Using (41) and (42) we see that $I_{s,\nu}$ is the following function of χ

$$I_{s,\nu}(\chi) = \left(\frac{\sqrt{M^2 + m^2 + 2Mm \cos \chi}}{M + m}\right)^s \cdot P_\nu\left(\frac{m + M \cos \chi}{\sqrt{M^2 + m^2 + 2Mm \cos \chi}}\right) \quad (48a)$$

which, for the particular case of equal masses, takes the simple form

$$I_{s,\nu}(\chi) = \cos^s \frac{1}{2}\chi P_\nu(\cos \frac{1}{2}\chi) \quad (48b)$$

With this definition the integrated equation (47) reads

$$\begin{aligned} \int_0^\infty \left\langle \frac{1 - I_{s,\nu}(\chi)}{a\tau(c)} \right\rangle h_\nu(c) c^{s+2} dc &= \frac{\nu(\nu + s + 1)}{2\nu - 1} \int_0^\infty h_{\nu-1}(c) c^{s+1} dc \\ &+ \frac{(\nu + 1)(s - \nu)}{2\nu + 3} \int_0^\infty h_{\nu+1}(c) c^{s+1} dc \end{aligned}$$

or in terms of averages

$$\begin{aligned} (2\nu + 1) \left\langle \frac{1 - I_{s,\nu}(\chi)}{a\tau(c)} c^s P_\nu(\cos \vartheta) \right\rangle \\ = \nu(\nu + s + 1) \langle c^{s-1} P_{\nu-1}(\cos \vartheta) \rangle \\ + (\nu + 1)(s - \nu) \langle c^{s-1} P_{\nu+1}(\cos \vartheta) \rangle \end{aligned} \quad (49)$$

I believe that equation (49) contains all possible derivable relations between averages as special cases. Some of the most notable ones are

listed below

$$s = 1, \quad \nu = 1$$

$$\left\langle \frac{1 - \cos \chi}{a\tau(c)} c \cos \vartheta \right\rangle = \frac{M + m}{M} \quad (50a)$$

$$s = 2, \quad \nu = 0$$

$$\left\langle \frac{1 - \cos \chi}{a\tau(c)} c^2 \right\rangle = \frac{(M + m)^2}{Mm} \langle c \cos \vartheta \rangle \quad (50b)$$

$$s = 2, \quad \nu = 2$$

$$\left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau(c)} c^2 P_2(\cos \vartheta) \right\rangle =$$

$$= \frac{4(M + m)^2}{M} \langle c \cos \vartheta \rangle \quad (50c)$$

While the averages entering into (49) are not always the desired ones, it remains true nevertheless that all solution methods evolved in the following use this equation system as a starting point rather than other forms of the Boltzmann equation.

IIB. THE MEAN FREE TIME MODEL AT HIGH FIELD

If the angular distribution in the center of mass system is independent of speed and the collision cross section varies inversely as the speed then the developments of the previous section permit actually a solution of the Boltzmann equation. It is a solution in the sense that all significant velocity averages can be obtained directly without the knowledge of the velocity distribution function.

Before developing these facts from the equations of the last section, I should point out that the derivation to follow is in a sense artificial. It has been shown already by Maxwell¹⁹ for related problems that if the mean free time between collisions is assumed constant specially simple techniques may be employed to get constants of experimental importance. These techniques can be employed here; they consist essentially in multiplying (13) by a suitable multiplier, followed by integration over $\mathbf{c}$. However, if we were to follow this procedure we would have to duplicate for a special model in an unsystematic way the work done systematically for all laws of interactions in the preceding section. A further advantage of using systematic procedure is that we can see at a

¹⁹ Maxwell, J. C., Collected Papers, Vol. II, p. 40.

glance what averages can or cannot be obtained, and what the relationship is between the high field and the general averages.

For this reason we limit ourselves at present to the high field averages obtainable from (49). For the special case under discussion this equation system takes the form

$$\begin{aligned} (2\nu + 1) \left\langle \frac{1 - I_{s,\nu}(\chi)}{a\tau} \right\rangle \langle c^s P_\nu (\cos \vartheta) \rangle \\ = \nu(\nu + s + 1) \langle c^{s-1} P_{\nu-1} (\cos \vartheta) \rangle \\ + (\nu + 1)(s - \nu) \langle c^{s-1} P_{\nu+1} (\cos \vartheta) \rangle \end{aligned} \quad (51)$$

that is we have a system of linear relations connecting the averages $\langle c^s P_\nu (\cos \vartheta) \rangle$. The connection between these averages is made apparent in Fig. 9. Each average $\langle c^s P_\nu (\cos \vartheta) \rangle$ is marked in this figure as a dot in an s - ν -plane if s is integer. The equations (51) connecting these averages are shown as lines with different equations leading to the same dot shown in different outline. These equations generally have the shape of a V ;

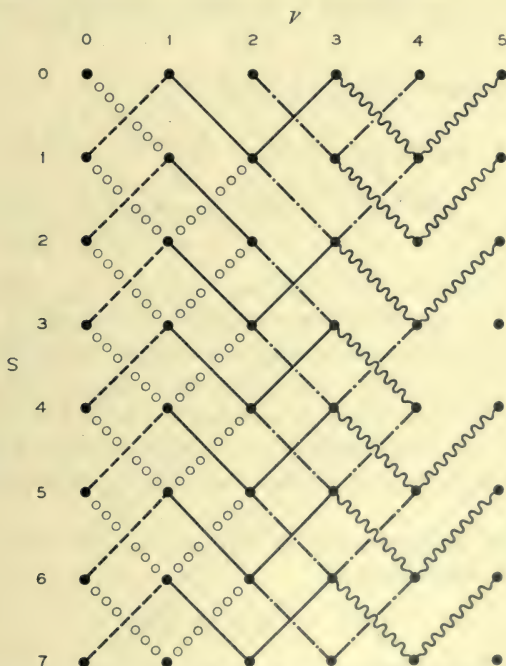


Fig. 9 — Interconnection established by the Boltzmann equation among the averages $\langle c^s P_\nu (\cos \vartheta) \rangle$; case of constant mean free time.

there are two notable exceptions to this rule, however, which make the recurrence method possible, the equations $\nu = 0$ have no left leg and the equations $s = \nu$ have no right leg. Starting out with the average $s = 0$, $\nu = 0$, which equals unity by definition one can thus proceed systematically as shown in Fig. 10, to get other averages. The averages reached are the ones for which s and ν are non-negative integers of equal parity with the restriction $s \geq \nu$. One verifies easily that this set is equivalent to the set of all products of integer powers of the velocity components.

The first three relations one uses in the path outlined in Fig. 10, are the simplified forms of the three equations (50). We find

$$\langle c_z \rangle = \frac{M + m}{M} \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle \quad (52)$$

$$\langle c^2 \rangle = \frac{(M + m)^3}{M^2 m} \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle^2 \quad (53)$$

$$\langle c^2 P_2(\cos \vartheta) \rangle = \frac{4(M + m)^3}{M^2 \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle}$$

or, more conveniently with the help of (53)

$$\langle c_z^2 \rangle = \frac{(M + m)^3 \left\langle \frac{M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle}{M^2 m \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle^2} \quad (54)$$

The three equations (52), (53) and (54) give the drift velocity, the total energy, and the energy partition of the travelling ion. Equation (52) gives a constant mobility and can actually be derived from a low field theory. Formula (52) thus states that for problems involving a constant mean free time the high field and low field mobilities are numerically identical. One would suspect that the intermediate field value would have to fall in line too. This is indeed the case as will be shown in Section IIIA.

A convenient interpretation of (53) may be had by combining (52) and (53) in the following way

$$\langle mc^2 \rangle = m \langle c_z \rangle^2 + M \langle c_z \rangle^2 \quad (55)$$

The left side is essentially the total energy of the ion, the first term on the right is the energy visible in the drift motion; it follows therefore

that the second term is the "invisible" or random part of the mean energy. Formula (55) thus states that

$$\frac{\text{random energy}}{\text{visible energy}} = \frac{\text{molecular mass}}{\text{ion mass}} \quad (56)$$

that is, it exhibits in a quantitative way the capacity of storing energy in the form of random motion which light ions travelling in a heavy gas possess; for ions travelling in the parent gas the ordered and the random part of the energy are just equal; for heavy ions in a light gas the disordered fraction becomes negligible.

There are various ways of understanding the implications of equation (54). One way is to derive the mean energy in a direction at right angles to the field by the use of (53). We find

$$\langle c_x^2 \rangle = \frac{(M + m)^3 \left\langle \frac{\sin^2 \chi}{a\tau} \right\rangle}{Mm \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle^2} \quad (57)$$

Now from (54) and (57) the partition of the energy e may be obtained.

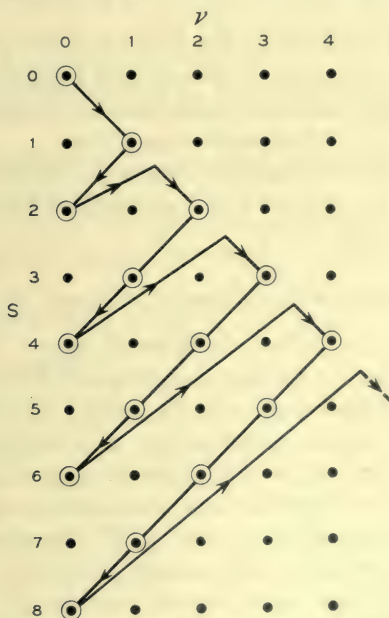


Fig. 10 — Order to be followed in computing by recursion the averages $\langle c^2 P, (\cos \vartheta) \rangle$; case of constant mean free time.

It comes out to be

$$e_x:e_y:e_z = \left\langle \frac{M \sin^2 \chi}{\tau} \right\rangle : \left\langle \frac{M \sin^2 \chi}{\tau} \right\rangle : \left\langle \frac{M \sin^2 \chi + 4m(1 - \cos \chi)}{\tau} \right\rangle \quad (58)$$

This shows up immediately the equipartition property for small m/M and the overwhelming preponderance of the motion in the field direction for large m/M . As an analysis of the random motion, however, equation (58) is deficient because the three directions become only comparable after the square of the drift velocity (52) is subtracted out of the z -component. We find

$$\langle c_z^2 \rangle - \langle c_z \rangle^2 = \frac{(M + m)^2 \left\langle \frac{M \sin^2 \chi + 2m(1 - \cos \chi)^2}{a\tau} \right\rangle}{Mm \left\langle \frac{3M \sin \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle^2} \quad (59)$$

and from this a more refined partition formula which only counts random motion

$$e_x:e_y:e_z^* = \left\langle \frac{\sin^2 \chi}{\tau} \right\rangle : \left\langle \frac{\sin^2 \chi}{\tau} \right\rangle : \left\langle \frac{M \sin^2 \chi + 2m(1 - \cos \chi)^2}{(M + m)\tau} \right\rangle \quad (60)$$

For small m/M this result does not differ essentially from (58) but if m/M is large the z component of the random energy does not grow indefinitely the way the total energy does. Instead it stops at a value which is about four times one of the other two values.

A discussion of these expressions for special models will be delayed until the equations are extended to intermediate field conditions. This will be done in Part III.

IIC. THE CASE OF LARGE MASS RATIOS: ELECTRONS OR HEAVY IONS

The distribution of velocities for a small value of m/M is treated in the literature because it applies to electrons.^{4,5} However, for the sake of completeness the derivation will be carried out here for the high field case. In this derivation all features of the law of scattering are left open, except that conservation of the kinetic energy is assumed.

The development in spherical harmonics carried out in Section IIA is the suitable starting point for small m/M , because, in this case, the distribution is almost spherically symmetrical and the expansion in spherical harmonics is also an expansion in powers of m/M . If we keep only $h_0(c)$ and $h_1(c)$ in the system (47) and treat m/M as small we get two equations, one for $\nu = 0$ and the one for $\nu = 1$. We may then use

(41) and (42) in the simplified form

$$\frac{c'}{c} \sim 1 + \frac{m}{M} (1 - \cos \chi) \quad (61)$$

$$\cos \kappa \sim \cos \chi + \frac{m}{M} \sin^2 \chi \quad (62)$$

and develop the equations in powers of m/M . Starting out with the simpler equation $\nu = 1$ we find

$$\left\langle \frac{1 - \cos \chi}{a\tau(c)} \right\rangle h_1(c) = -\frac{dh_0(c)}{dc} \quad (63)$$

This same procedure is not adequate for the equation $\nu = 0$ because the two left hand terms in (47) cancel in zero approximation. We must therefore develop the integral up to linear terms in m/M ; this is a perfectly straightforward, though somewhat cumbersome, step. It leads to the following equation

$$\begin{aligned} \frac{m}{M} \left[c \frac{d}{dc} \left\langle \frac{1 - \cos \chi}{a\tau(c)} \right\rangle h_0(c) \right] + 3 \left\langle \frac{1 - \cos \chi}{a\tau(c)} \right\rangle h_0(c) \\ = \frac{1}{3} \left[\frac{dh_1(c)}{dc} + \frac{2}{c} h_1(c) \right] \end{aligned}$$

The equation may be integrated by multiplication with c^2 ; this yields

$$h_1(c) = 3 \frac{m}{M} \left\langle \frac{1 - \cos \chi}{a\tau(c)} \right\rangle c h_0(c) \quad (64)$$

Elimination of $h_1(c)$ from (63) and (64) gives a differential equation for $h_0(c)$ which is easily solved by quadrature; the result is

$$h_0(c) = \exp \left[-3 \frac{m}{M} \int_0^c \left\langle \frac{1 - \cos \chi}{a\tau(c)} \right\rangle^2 c \, dc \right] \quad (65)$$

Except for the dependence on the angular law of scattering this formula may be found in the literature.⁴ Its most important special cases are obtained for $\tau = \text{const}$ (Pseudo-Maxwellian distribution) and $\tau = \text{const}/c$ (Druyvesteyn distribution).

The derivation of (65) should be completed by a proof that indeed $h_2(c)$ is small compared to $h_1(c)$. This is not true for all values of c ; on the contrary, the argument below shows that near the origin where $c \sim a\tau(c)$, $h_2(c)$ is actually comparable to $h_1(c)$. For our purposes, however, it is sufficient if it is true in the overwhelming majority of cases. As the proof applies equally to all h_i 's we will run it in this manner. Assuming

provisionally that

$$h_{\nu+1} \ll h_{\nu} \ll h_{\nu-1}$$

we see that we can drop the terms in $h_{\nu+1}$ in the system (47). The integral in (47) is evaluated in zero order with the help of (61) and (62); it becomes then

$$\frac{h_{\nu}(c)}{\tau(c)} \cdot \frac{1}{2} \int_0^{\pi} P_{\nu}(\cos \chi) \Pi(\chi) \sin \chi d\chi$$

This means that we may solve explicitly for $h_{\nu}(c)$ in terms of $h_{\nu-1}(c)$. The formula is

$$h_{\nu}(c) \approx \frac{\nu}{2\nu-1} \frac{-\frac{dh_{\nu-1}(c)}{dc} + \frac{\nu-1}{c} h_{\nu-1}(c)}{\left\langle \frac{1 - P_{\nu}(\cos \chi)}{a\tau(c)} \right\rangle} \quad (66)$$

To estimate the order of magnitude of this we may neglect $P_{\nu}(\cos \chi)$ as compared to 1 and assume ν large. The operation in the numerator will lead to two kinds of terms: some of the form

$$\frac{1}{c} h_{\nu-1}(c)$$

and others of the type

$$\frac{m}{M} \frac{c}{\{(a\tau(c))\}^2} h_{\nu-1}(c)$$

coming from differentiation of an exponent of the type (65). Now we find from (65) that the overwhelming majority of particles have speeds c which in order of magnitude satisfy

$$c \sim (M/m)^{1/2} a\tau(c) \quad (67)$$

because this is the range within which the exponent remains comparable to 1. Applying this to the two types of terms arising from (66) we find for them in order of magnitude

$$\nu a\tau(c) \frac{1}{c} h_{\nu-1}(c) \sim \nu \left(\frac{m}{M}\right)^{1/2} h_{\nu-1}(c)$$

and

$$a\tau(c) \frac{m}{M} \frac{c}{\{(a\tau(c))\}^2} h_{\nu-1}(c) \sim \left(\frac{m}{M}\right)^{1/2} h_{\nu-1}(c)$$

If we substitute this into (66) we see that the h_ν 's decrease as $(m/M)^{\frac{1}{2}\nu}$, with a possible $\nu!$ slowing up the final convergence. In any case $h_2(c)$ comes out small compared to $h_1(c)$ which is all that is needed to make equation (65) approximately correct.

While the case of small m/M is generally known it appears to be otherwise for large m/M . The intuitive basis for the solution of this case is the fact observable from (52), (57) and (59) that $\langle c \rangle$ increases indefinitely with m/M , but that the relative deviation from the mean decreases so that the distribution function approaches a δ -function. The structure of this limiting function may be explored, starting directly from (40), because in the limit of large m/M the sphere of integration shrinks as may be verified from (37). This makes it possible to replace the integral in (40) by differential terms. This becomes clearer if (40) is written in the form

$$a \frac{\partial h(c)}{\partial c_z} = \frac{1}{2} \int_0^\pi \Pi(\chi) \sin \chi \, d\chi \frac{1}{2\pi} \int_0^{2\pi} d\omega \left\{ \left(\frac{c'}{c} \right)^3 \frac{h(c')}{\tau(c')} - \frac{h(c)}{\tau(c)} \right\} \quad (68)$$

Let us call the inner average $\mathcal{J}$. It exhibits the differential properties discussed earlier: the curly bracket is the difference of two terms which are almost identical. Hence we approximate the value of $\mathcal{J}$ by expanding the slowly varying terms to first order in $c' - c$, while the fast varying $h(c')$ will be expanded to square terms. This expansion is obviously permissible for everything except the rapidly varying function $h(c')$. For $h(c')$ itself no justification can be offered except success. By proceeding to square terms in this expansion we mitigate any possible error committed, but it is quite possible that structural details are lost in the procedure.

The development of $\mathcal{J}$ is straightforward. We proceed as follows

$$\begin{aligned} \mathcal{J} &= \frac{1}{2\pi} \int_0^{2\pi} d\omega \left\{ h(c) \left(\left(\frac{c'}{c} \right)^3 \frac{1}{\tau(c')} - \frac{1}{\tau(c)} \right) + \left(\frac{c'}{c} \right)^3 \frac{1}{\tau(c')} (h(c') - h(c)) \right\} \\ &\approx h(c) \left\{ \left(\frac{c'}{c} \right)^3 \frac{1}{\tau(c')} - \frac{1}{\tau(c)} \right\} + \frac{1}{\tau(c)} \cdot \frac{1}{2\pi} \int_0^{2\pi} d\omega \{ h(c') - h(c) \} \end{aligned}$$

The expansion of the first term involves only (41) which to this order reads

$$c' - c \approx c \frac{M}{m} (1 - \cos \chi)$$

This formula does not contain the azimuth ω which therefore disappears trivially. In the second term on the contrary we are dealing with a vectorial difference involving all three polar coordinates c' , κ , ω of c' .

If we set

$$c'_\xi = c' \sin \kappa \cos \omega$$

$$c'_\eta = c' \sin \kappa \sin \omega$$

$$c'_\zeta = c' \cos \kappa$$

we find

$$\begin{aligned} h(c') - h(c) &= \frac{\partial h(c)}{\partial c_\xi} c'_\xi + \frac{\partial h(c)}{\partial c_\eta} c'_\eta + \frac{\partial h(c)}{\partial c_\zeta} (c'_\zeta - c) + \frac{1}{2} \frac{\partial^2 h(c)}{\partial c_\xi^2} c'^2_\xi \\ &+ \frac{1}{2} \frac{\partial^2 h(c)}{\partial c_\eta^2} c'^2_\eta + \frac{1}{2} \frac{\partial^2 h(c)}{\partial c_\zeta^2} (c'_\zeta - c)^2 + \frac{\partial^2 h(c)}{\partial c_\xi \partial c_\eta} c'_\xi c'_\eta \\ &+ \frac{\partial^2 h(c)}{\partial c_\xi \partial c_\zeta} c'_\xi (c'_\zeta - c) + \frac{\partial^2 h(c)}{\partial c_\eta \partial c_\zeta} c'_\eta (c'_\zeta - c) + \dots \end{aligned}$$

With the formulas given, integration over ω is elementary. We find

$$\begin{aligned} \mathcal{J} &\approx h(c) \left\{ \left(\frac{c'}{c} \right)^3 \frac{1}{\tau(c')} - \frac{1}{\tau(c)} \right\} + \frac{1}{\tau(c)} \left\{ \frac{\partial h(c)}{\partial c_\zeta} (c' \cos \kappa - c) \right. \\ &\left. + \frac{1}{4\tau(c)} \left(\frac{\partial^2 h(c)}{\partial c_\xi^2} + \frac{\partial^2 h(c)}{\partial c_\eta^2} \right) c'^2 \sin^2 \kappa + \frac{1}{2\tau(c)} \frac{\partial^2 h(c)}{\partial c_\zeta^2} (c' \cos \kappa - c)^2 \right\} \end{aligned}$$

All coefficients are to be evaluated only to the lowest non-vanishing order in M/m . From the equations (41) and (42) we get

$$c' \cos \kappa - c \approx (M/m) c(1 - \cos \chi)$$

$$c'^2 \sin^2 \kappa \approx (M/m)^2 c^2 \sin^2 \chi$$

The first term in $\mathcal{J}$ is simplified further by introduction of the parameter α used in the dimensional analysis of Section ID. According to that section we have that

$$\frac{d \ln \tau(c)}{d \ln c} = -1 + \alpha \quad (69)$$

We can generalize the original definition for any $\tau(c)$ by the above equation, where α is now a function of c . Eliminating also the tempo-

rary device of a ξ, η, ζ coordinate system we find for $\mathcal{G}$:

$$\begin{aligned} \mathcal{G} = & \frac{M}{m} (4 - \alpha(c))(1 - \cos \chi) \frac{h(c)}{\tau(c)} + \frac{M}{m} \frac{(1 - \cos \chi)}{\tau(c)} \left\{ c_x \frac{\partial h}{\partial c_x} + c_y \frac{\partial h}{\partial c_y} \right. \\ & \left. + c_z \frac{\partial h}{\partial c_z} \right\} + \left(\frac{M}{m} \right)^2 \frac{\sin^2 \chi}{4\tau(c)} \left\{ \frac{\partial^2 h}{\partial c_x^2} + \frac{\partial^2 h}{\partial c_y^2} + \frac{\partial^2 h}{\partial c_z^2} \right\} \\ & + \left(\frac{M}{m} \right)^2 \frac{2(1 - \cos \chi)^2 - \sin^2 \chi}{4\tau(c)} c \left\{ c_x \frac{\partial}{\partial c_x} + c_y \frac{\partial}{\partial c_y} + c_z \frac{\partial}{\partial c_z} \right\} \\ & \cdot \left\{ \frac{c_x}{c} \frac{\partial h}{\partial c_x} + \frac{c_y}{c} \frac{\partial h}{\partial c_y} + \frac{c_z}{c} \frac{\partial h}{\partial c_z} \right\} \end{aligned}$$

If the last term is evaluated exactly terms of the order $(M/m)^2$ are added to the first derivative terms in $h(c)$. They are obviously negligible. Finally integration over χ yields the following form for equation (68)

$$\begin{aligned} a \frac{\partial h(c)}{\partial c_z} = \langle \mathcal{G} \rangle = & \frac{M}{m} \left\langle \frac{1 - \cos \chi}{\tau(c)} \right\rangle \{4 - \alpha(c)\} h(c) \\ & + \frac{M}{m} \left\langle \frac{1 - \cos \chi}{\tau(c)} \right\rangle \sum_{i=1}^3 c_i \frac{\partial h}{\partial c_i} + \frac{1}{4} \left(\frac{M}{m} \right)^2 \left\langle \frac{\sin^2 \chi}{\tau(c)} \right\rangle c^2 \sum_{i=1}^3 \frac{\partial^2 h}{\partial c_i^2} \quad (70) \\ & + \frac{1}{4} \left(\frac{M}{m} \right) \left\langle \frac{(1 - \cos \chi)(1 - 3 \cos \chi)}{\tau(c)} \right\rangle \sum_{i,k=1}^3 c_i c_k \frac{\partial^2 h}{\partial c_i \partial c_k} \end{aligned}$$

When equation (70) is considered up to linear terms in M/m it yields a δ -function about the drift velocity $\langle c_z \rangle$ which results from the implicit equation

$$\langle c_z \rangle = \frac{m}{M} a \left/ \left\langle \frac{1 - \cos \chi}{\tau(c)} \right\rangle \right|_{c=c_z} \quad (71)$$

The δ -function takes here the aspect of a non integrable function which in a special case can be seen to equal

$$h \approx \{c_x^2 + c_y^2 + (c_z - \langle c_z \rangle)^2\}^{-3/2}$$

When normalization is imposed on such a function it is made to vanish everywhere except at the point corresponding to the drift velocity.

The square terms in M/m are necessary to gain information about the functional form of $h(c)$. Since the region in velocity space in which h is appreciable is still small we may take α as constant in that region. A further simplification results from order of magnitude considerations on c :

$$c^2 \sim c_x^2 \sim \langle c_x \rangle^2 \gg c_x^2 \sim c_y^2 \sim (c_z - \langle c_z \rangle)^2$$

In working out the details we see that division by

$$\frac{M}{m} \left\langle \frac{1 - \cos \chi}{\tau(\mathbf{c})} \right\rangle$$

puts the first two terms on the right hand side of (70) in a simple form. The coefficient of the field term becomes then

$$\frac{am}{M \left\langle \frac{1 - \cos \chi}{\tau(\mathbf{c})} \right\rangle} = \frac{am}{M \left\langle \frac{1 - \cos \chi}{\tau(\mathbf{c})} \right\rangle_{c=\langle c_z \rangle}} \cdot \left(\frac{\langle c_z \rangle}{c} \right)^{1-\alpha}$$

The first factor is just $\langle c_z \rangle$ by the identity (71). For the second factor we have up to linear small terms

$$c = \sqrt{c_x^2 + c_y^2 + c_z^2} = \sqrt{c_x^2 + c_y^2 + (\langle c_z \rangle + c_z - \langle c_z \rangle)^2} \\ \approx \langle c_z \rangle \left\{ 1 + \frac{c_z - \langle c_z \rangle}{\langle c_z \rangle} \right\}$$

and hence for the coefficient of the field term

$$\langle c_z \rangle \left(\frac{\langle c_z \rangle}{c} \right)^{1-\alpha} \approx \langle c_z \rangle - (1 - \alpha)(c_z - \langle c_z \rangle)$$

After division by

$$\frac{M}{m} \left\langle \frac{1 - \cos \chi}{\tau(\mathbf{c})} \right\rangle,$$

the square terms still contain another small factor M/m ; it appears sufficient, therefore, to keep only the leading terms which are the ones containing $\langle c_z \rangle^2$ as factor. All these terms are multiplied with the ratio of two angular averages over $d\chi$; these may be taken as independent of c to a good approximation. Equation (70) thus takes the form

$$(4 - \alpha)h(\mathbf{c}) + c_x \frac{\partial h(\mathbf{c})}{\partial c_x} + c_y \frac{\partial h(\mathbf{c})}{\partial c_y} + (2 - \alpha)(c_z - \langle c_z \rangle) \frac{\partial h(\mathbf{c})}{\partial c_z} \\ + \frac{1}{4} \frac{M}{m} \frac{\left\langle \frac{\sin^2 \chi}{\tau} \right\rangle}{\left\langle \frac{1 - \cos \chi}{\tau} \right\rangle} \langle c_z \rangle^2 \left\{ \frac{\partial^2 h(\mathbf{c})}{\partial c_x^2} + \frac{\partial^2 h(\mathbf{c})}{\partial c_y^2} \right\} \\ + \frac{1}{2} \frac{M}{m} \frac{\left\langle \frac{(1 - \cos \chi)^2}{\tau} \right\rangle}{\left\langle \frac{1 - \cos \chi}{\tau} \right\rangle} \langle c_z \rangle^2 \frac{\partial^2 h(\mathbf{c})}{\partial c_z^2} = 0 \quad (72)$$

The equation can be solved explicitly in Cartesian coordinates by the method of separation of variables. The result is

$$h(\mathbf{c}) = \exp \left[-2 \frac{m}{M} \frac{\langle \frac{1 - \cos \chi}{\tau} \rangle}{\langle \frac{\sin^2 \chi}{\tau} \rangle} \frac{c_x^2 + c_y^2}{\langle c_z \rangle^2} - (2 - \alpha) \frac{m}{M} \frac{\langle \frac{1 - \cos \chi}{\tau} \rangle}{\langle \frac{(1 - \cos \chi)^2}{\tau} \rangle} \frac{(c_z - \langle c_z \rangle)^2}{\langle c_z \rangle^2} \right] \quad (73)$$

This is a Maxwellian distribution with elliptic distortion and shifted origin, that is, the type shown in Fig. 1 (b).

The result (73) indicates the main features of the solution for heavy ions. Because of the neglect of derivatives higher than the second in $h(\mathbf{c}')$ it is not certain that (73) is correct in all details, even in the limit of very large m/M .

III. THE CASE OF EQUAL MASSES; IONS TRAVELLING IN THE PARENT GAS

The developments of the previous section show that if the ion mass is either large or small in comparison to the molecular mass, analytical methods can be applied successfully to determine the velocity distribution of the ions. No such possibility was found for the mass ratio unity, which one would judge to be of particular interest because it applies to ions travelling in the parent gas. There exist isolated fragments of such an analytical theory; for instance, if we assume isotropic scattering in (46), that is $\Pi(\chi) = 1$, then the zeroth equation (46) becomes explicitly integrable and yields

$$h_0(c) = -\frac{1}{3} a\tau(c) \frac{dh_1(c)}{dc} \quad (74)$$

This is a curious reversal of the differential relationship (63) derived for electrons and implies a rather strong condition on the structure of $h_1(c)$. However, I have not been able to consolidate these fragments into something which can be used successfully in computation. The high field distribution function for mass ratio unity appears, however, sufficiently interesting to warrant the use of other methods.

A numerical determination of the velocity distribution function was undertaken in cooperation with R. W. Hamming by the so-called Monte Carlo method. The Monte Carlo method is a way of gaining

statistical information about a system by following an individual member through a large number of random processes. The result of such a procedure is knowledge about one member of the assembly for a long period of time. Time averages of various kinds can be obtained from such data; these time averages are then set equal to instantaneous averages over the assembly, in accordance with ergodic theory. In our case, an ion was followed through 10,000 collisions. On an average, the collisions were isotropic in the center of mass system ($\Pi(\chi) = 1$) and obeyed a mean free time condition $\tau = \text{const.}$ Actually, both the free time and the scattering angles varied from collision to collision; the angles varied in a random fashion over a unit sphere and τ was random within an exponential distribution.

A Monte Carlo calculation of this type consists of three parts. In the first part the random numbers having the required distributions are obtained and recorded. In the present problem there were three such random numbers required for each collision, namely a time and two angles. These numbers were placed on 10,000 IBM cards, along with suitable identification. In the second part a calculating machine simulates the successive collisions and keeps a record of the initial and final velocities for each one. The third part consists in analyzing statistically the numerical material accumulated in the second. For the first part of the calculation particular values must be chosen for the acceleration a and the mean free time τ . These values were

$$a = 1$$

$$\tau = \log_{10} e = 0.43429$$

However, the dimensional analysis of Section ID shows us at this point that these two constants enter into the problem only through their product $a\tau$ which scales all velocities. It is therefore convenient at the statistical stage to remove these factors and to analyze the results in terms of a dimensionless variable which by (26c) we take in the form

$$w = \frac{c}{a\tau} \quad (75)$$

In view of the a priori information for mean free time problems which is gathered in Section IIB we can use the statistical data from the Monte Carlo calculation in two ways. We may (a) check the numerical computation itself or (b) gain new information not available otherwise.

(a) The check of the numerical calculation by statistical analysis proceeds as follows. From deductive reasoning we have obtained the

averages (52), (54) and (57) for c_x , c_x^2 and c_x^3 . These formulas ought to be verified in the Monte Carlo calculation. This is indeed approximately true. A sampling covering 9492 out of the 10,000 collisions gives

	by Monte Carlo	by deduction
$\langle w_x \rangle$	1.912	2
$\langle w_x^2 \rangle$	0.801	$\frac{8}{9} = 0.889$
$\langle w_x^3 \rangle$	5.165	$\frac{56}{9} = 6.222$

The agreement is essentially there but deviations are noticeable. In judging these we have to realize that fluctuations are quite large in this problem. For instance if the calculation is broken down into ten runs of approximately 1,000 events each one finds the following time averages for the partial runs:

$\langle w_x \rangle$	$\langle w_x^2 \rangle$	$\langle w_x^3 \rangle$	
1.821	0.837	4.453	
1.975	0.748	5.531	
1.915	0.785	5.132	
1.954	0.829	5.441	
1.868	0.731	4.794	
2.003	0.807	5.581	
1.766	0.793	4.811	
1.962	0.827	5.360	
2.003	0.901	5.471	
1.914	0.727	5.200	
predicted	2.000	0.889	6.222

Among these runs there are some having averages higher than the predicted values, but the data clearly show that the Monte Carlo averages are generally lower. In the search for reasons it was first felt that perhaps the desired mean value for τ is not actually reached, perhaps through systematic errors introduced by the operator when rejecting certain runs. This seems indeed to be the case. The mean free time obtained from the 9492 runs mentioned above is

$$\tau = 0.4269$$

which is slightly low. Indeed it is observed that the runs with high τ were particularly troublesome in the calculation and were preferably rejected by the operator. It seems doubtful however that this error

could account for the entire discrepancy, particularly in the mean squares, although it must be emphasized that the runs with high τ make a more than proportional contribution to the total average. The angles of scattering have not been subjected to a similar analysis so that we cannot make a statement whether the aimed at isotropy in the law of scattering was realized or not. We conclude therefore by saying that while the Monte Carlo calculation gives results in general agreement with the deductive theory there are small but noticeable systematic errors in it whose origin is only partly explained. Similar errors must exist in the new results which cannot be compared with theoretical predictions.

(b) In this part we will discuss the velocity distribution function which may be constructed from the Monte Carlo results. In constructing such a function we make use of the fact that, between collisions, the velocity is accelerated at a uniform rate. Thus, in each period between two collisions, the velocity vector traces out a straight line parallel to the w_z axis covering equal distances in equal times. The Monte Carlo calculation furnishes us with a number of such straight lines as shown in Fig. 11. The density of these straight line tracks in velocity space is the velocity distribution function. The actual procedure used to obtain it was to lay a grid with a mesh of 0.23 in a half-plane with coordinates w_z and $w_\rho = \sqrt{w_x^2 + w_y^2}$ and to count the number of lines crossing each horizontal square edge. When the resultant count is converted to density

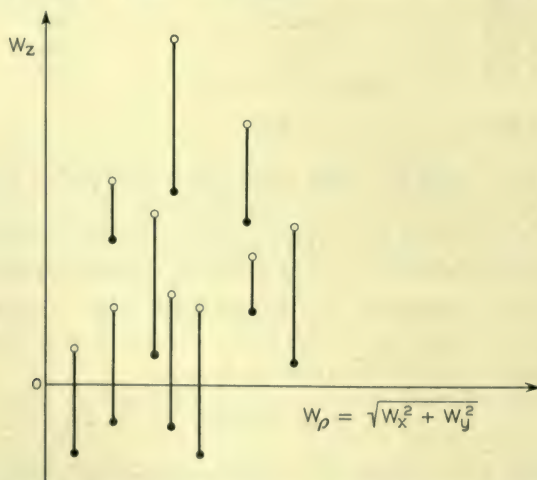


Fig. 11 — Straight line pattern in the $w_z - w_\rho$ half plane from which the velocity distribution is constructed; the Monte Carlo calculation furnishes the initial and final velocities (dots and rings).

and normalized to 1 we get a distribution in the $w_z - w_p$ half plane which is shown in Fig. 12. Division by $2\pi w_z$ will transform it into a conventional distribution function in velocity space; a plot of this function is shown in Fig. 13. What distinguishes this distribution function from functions previously proposed is the elongated probability contours. This feature is not unexpected in view of the unequal energy partition apparent in the equations (58) and (60).

The probability contours shown in Figs. 12 and 13 give a reliable general picture but we must not expect from them fine detail. Indeed we will now prove that the distribution function is infinite along the entire positive c_z -axis, a feature which is not obvious from inspection of Fig. 13.

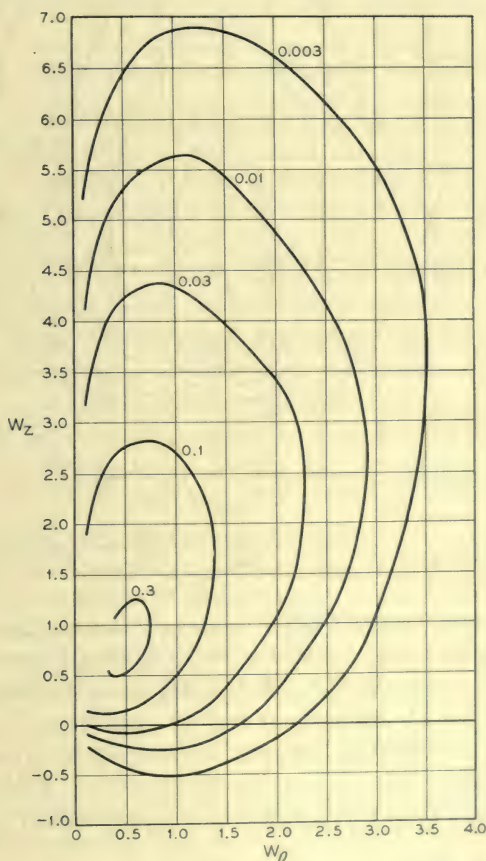


Fig. 12 — Motion of ions through the parent gas in a high field; distribution of velocities in the $w_z - w_p$ half plane resulting from the Monte Carlo calculation.

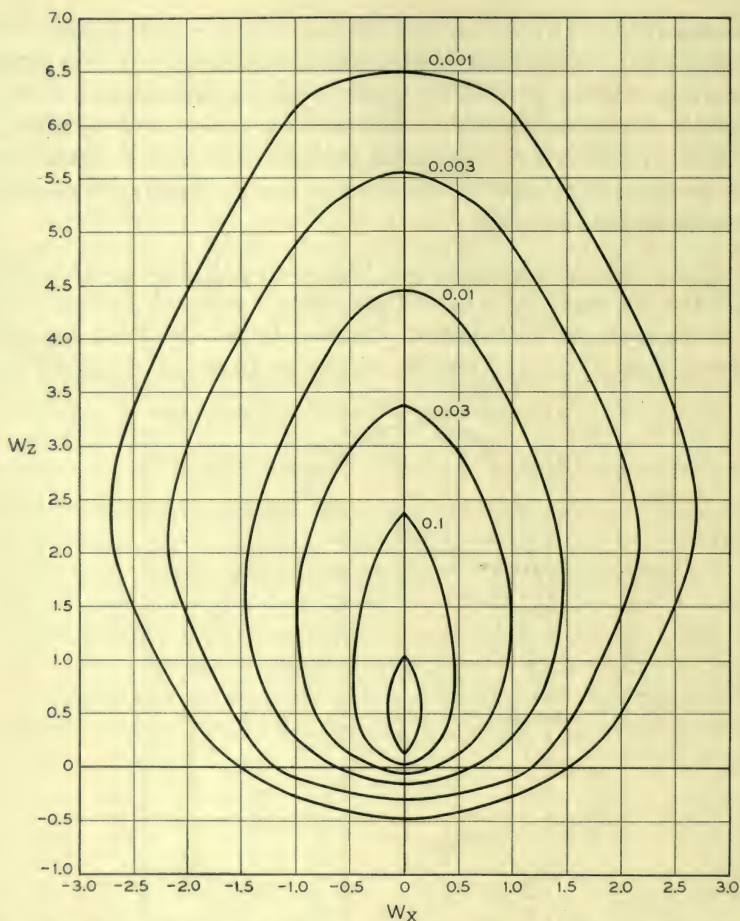


Fig. 13 — Velocity distribution function of ions moving through the parent gas in a high field; contours constructed from the Monte Carlo results of Fig. 12.

A simple physical proof of this statement goes as follows. Suppose an ion and a molecule make a collision which is almost central, but has a small impact parameter b . The collision will bring the ion almost to rest because the atom was originally at rest by hypothesis. Because the collision was not quite central, however, the ion will have a small residual velocity c_f at right angles to its original velocity c_i . For any reasonable law of scattering this quantity c_f will be proportional to c_i and to b .

$$c_f \propto c_i \cdot b$$

The probability for a value of b between b and $b + db$ is proportional

to $b \, db$. Thus even if all c_i 's were equally probable the probability for c_f would vary as $c_f \, dc_f$. Actually very small c_i 's may be specially probable as the theorem states and this fact may or may not increase the probability for small c_f . This means that $P(c_f)dc_f$ probably varies as $c_f dc_f$, and may perhaps even contain a smaller power of c_f . When such a probability function is plotted in velocity space it will vary as $1/c_f$. Thus we know that the distribution function $\varphi(\mathbf{c})$ for ions immediately following a collision has a singularity at the origin at least as $1/c_f$. The actual distribution function $h(\mathbf{c})$ is derived from this one by spreading each point out in the forward direction as shown in Fig. 11. For the mean free time case we can write this out explicitly in the form

$$h(\mathbf{c}) = \frac{1}{\tau} \int_0^\infty \varphi(\mathbf{c} - a\mathbf{t}) e^{-\frac{t}{\tau}} dt \quad (76)$$

For the case of a mean free path or other laws the formula is more awkward but they all differ from the above only in replacing $e^{-\frac{t}{\tau}}$ by a more complicated weight function. The singularity of h arises out of the singularity of φ which contains at least a factor $1/\sqrt{c_x^2 + c_y^2 + (c_z - at)^2}$. Along the c_x -axis, this become a factor $1/(c_x - at)$; this factor makes the integral diverge for all positive c_x ; as we approach the origin from negative c_x 's the distribution function will become infinite at least as $\ln 1/c_x$.

The reasoning given is intrinsically classical because of the use of "infinitely small" impact parameters. We should not hasten to conclude, however, that the quantization of the angular momentum will necessarily remove the singularity. Indeed we know that the only mechanical information which has to be put into the Boltzmann equation (31) is the differential cross-section for scattering. If this quantity does not differ essentially in the 180° direction from a classical cross section then it will not modify the conclusion we have reached.

Conclusions which are more informative, but less "anschaulich" may be obtained from a study of Boltzmann's equation either in its closed form (34) or (40) or its "Legendre" form (46) or (47). In view of the proof given we will give only an outline of the reasoning. First we can remove the second term in (40) by the substitution

$$h(\mathbf{c}) = \exp \left[- \int_0^{c_x} \frac{dc_x}{a\tau(\bar{c})} \right] h^*(\mathbf{c})$$

The exponential is easily seen to be always positive and finite for finite c . The Boltzmann equation then takes the form

$$\exp \left[- \int_0^{c_x} \frac{dc_x}{a\tau(\bar{c})} \right] a \frac{\partial h^*}{\partial c_x} = \frac{1}{c} \cdot (\text{an integral})$$

For equal masses the integral on the right runs over a plane in velocity space. Its integrand is always positive; hence the integral can never vanish and is always positive; it is conceivable that it could be infinite for a special position provided the infinity is integrable. Such an infinity would only make matters worse. At any rate the equation shows that $h^*(\mathbf{c})$ increases monotonely with increasing c_z . When we approach the origin from negative c_z we get a logarithmic divergence (or worse). As the function $h^*(\mathbf{c})$ can nowhere decrease with increasing c_z the infinity along the positive c_z axis is confirmed and its logarithmic nature is made very likely.

Information obtainable from equation (46) confirms this conclusion. Lowest powers in the entire recursion system can be made to cancel by assuming that for small c

$$h_0 \sim -A \ln w$$

$$h_i \sim B_i \quad i > 0$$

with suitable relationships existing between these quantities.

A defect of all three approaches is that they give no information concerning the nature of the infinity for $c_z > 0$. One is tempted to conclude from Fig. 13 that it cannot be very strong. Something like a singularity is discernible at the origin, particularly if the contour 0.1 is drawn back to cut the w_z -axis at a negative value; this is perfectly compatible with the available information. For large positive w_z , on the other hand, the picture almost contradicts the theorem just proved. One concludes from this that the singularity, for large c_z , becomes a weak and narrow ridge rising more or less abruptly in an otherwise well behaved function.

II. THE CASE OF EQUAL MASSES; A NEW COMPUTATIONAL PROCEDURE

The foregoing sections have accumulated substantial evidence that there are many analytical details involved when one discusses the structure of a velocity distribution function. These details are of little interest to the experimenter who may want nothing but a formula for the drift velocity or the average energy. In view of this situation it appears very desirable to find a method whereby such quantities can be derived directly and accurately from the Boltzmann equation without a full knowledge of the entire distribution.

Maxwell's original work shows us how to achieve this for molecules obeying the mean free time condition of Section IIB. In the following, a general method is described which will permit determination of such

averages for an arbitrary law of force between the ions and the gas molecules, and an arbitrary mass ratio. The application will be limited to the case of the mass ratio 1 whose study was begun in the preceding section.

The basis of the method is an observation on the equation system (46) or (47), which is the form taken by the Boltzmann equation after inserting the Legendre decomposition (43). It would appear at first sight that these recursion relations are of such a structure that an arbitrary function $h_0(c)$ could be substituted into the "zeroth" equation and that the relations would then successively determine $h_1, h_2, h_3 \dots$. Upon closer inspection this is found not to be the case. Suppose we have obtained somehow functions $h_0, h_1, h_2 \dots h_n$ and we are trying to use the n th equation to determine h_{n+1} . This equation is of the form

$$\frac{dh_{n+1}}{dc} + \frac{n+2}{c} h_{n+1} = \text{known material} \quad (77)$$

We solve for h_{n+1} by multiplying with c^{n+2} and integrating. This gives

$$c^{n+2} h_{n+1}(c) = \int^c (\text{known material}) dc$$

The left-hand side is of such a structure that it must vanish both for $c = 0$ and $c = \infty$. It follows that the right-hand integral when taken between the limits 0 and ∞ must equal zero. This condition is indeed obeyed for any $h_0(c)$ when $n = 0$. The integral condition reads in this case

$$\frac{1}{2} \int_0^\infty c^2 dc \int_0^\pi \Pi(\chi) \sin \chi d\chi \frac{h_0(c')}{\tau(c')} \left(\frac{c'}{c}\right)^3 - \int_0^\infty \frac{h_0(c)}{\tau(c)} c^2 dc = 0$$

If we invert the order of integration in the double integral, then introduce c' as variable of integration by equation (41) and finally invert again this becomes

$$\int_0^\infty \frac{h_0(c)}{\tau(c)} c^2 dc \left[\frac{1}{2} \int_0^\pi \Pi(\chi) \sin \chi d\chi - 1 \right] = 0$$

This equation is trivially obeyed because the square bracket vanishes in virtue of the definition of $\Pi(\chi)$. For values of n higher than 0, the integrability condition deduced from (77) is not generally obeyed for any function $h_0(c)$. Such a statement may be proved by examples; these examples will arise in the course of the calculations to follow. Thus we find that except in the passage from $h_0(c)$ to $h_1(c)$, the recursion system is such that at each stage it imposes a condition upon the h_r 's already determined if the new $h_{n+1}(c)$ is to exist at all. With such an infinity of

conditions one can improve indefinitely an initial trial function assumed for $h_0(c)$.

The integrability conditions whose general structure is thus indicated have actually already been written down. They are the equations (49) for the special case $s = \nu$. Generally speaking, the relations (49) are also of the recursion type, permitting us to start with arbitrary averages $\langle c^s \rangle$, and computing successively $\langle c^s P_1(\cos \vartheta) \rangle$ etc. At every stage, however, there is the exception mentioned: the equation for which $s = \nu$ has no third member, and therefore it imposes a condition upon averages already known from the previous equations. We shall refer to this type of equation as a "truncated" relation.

It is reasonable to assume that $1/\tau(c)$ can be developed into a power series in c because it equals the known constant polarization value for $c = 0$. If this can be assumed then each truncated relation $s = \nu$ is equivalent to a unique relation among velocity averages involving $h_0(c)$ only. One obtains this relation by applying to each member in the truncated relation its own recursion formula and repeating this process until ν is brought down to zero. This process will never lead into another truncated relation $s' = \nu'$ because at each step s' increases by at least two units with respect to ν' .

In order to test the method for a known case, it will be applied first to the case of constant mean free time. This case is adequately described by the theoretical treatment of Section IIB and the Monte Carlo calculation of Section IID. We have seen that the equations (49) reduce in this case to the form (51) which dovetails as shown in Fig. 9; this dovetailing leads to explicit values for certain averages as shown in Fig. 10. A "computational method" is only needed when one tries to get an average outside this selected list. In the present case the reduction of the truncated relations to a condition on $h_0(w)$ is particularly simple as is seen from Fig. 9. A singular relation which starts out as between $\langle c^{\nu-1} P_{\nu-1}(\cos \vartheta) \rangle$ and $\langle c^\nu P_\nu(\cos \vartheta) \rangle$ actually yields the numerical value of the latter because the former has been obtained numerically in a previous stage. This numerical value yields in combination with previous information $\langle c^{\nu+1} P_{\nu-1}(\cos \vartheta) \rangle$, $\langle c^{\nu+2} P_{\nu-2}(\cos \vartheta) \rangle$ etc and finally $\langle c^{2\nu} \rangle$. Thus we end up with the set of even moments of $h_0(c)$ which may be used in succession to determine $h_0(c)$ more and more closely. There is no guarantee that this procedure converges mathematically, since the general theorems usually require the knowledge of all integer moments.²⁰

²⁰ Shohat, J. A., and J. D. Tamarkin, *The Problem of Moments*. Am. Math. Soc., 1943. The original three-dimensional formulation appears a little more favorable for a proof because, in this case, we know indeed all integer moments.

The justification for the method rests therefore on an empirical basis at this point.

Assuming isotropic scattering, as in the "Monte Carlo" calculation we express our results in terms of the dimensionless variable w defined in (75). The equation system (51) becomes then

$$\begin{aligned} (2\nu - 1)(1 - \langle I_{s,\nu}(\chi) \rangle) \langle s, \nu \rangle = \\ = \nu(\nu + s + 1) \langle s - 1, \nu - 1 \rangle + (\nu + 1)(\nu - s) \langle s - 1, \nu + 1 \rangle \end{aligned} \quad (78)$$

where the abbreviation $\langle s, \nu \rangle$ has been introduced for $\langle w^s P_\nu(\cos \vartheta) \rangle$ and the quantities $\langle I_{s,\nu}(\chi) \rangle$ are simple numbers computable from (48b) and the assumption of isotropic scattering. The first truncated relation is $s = \nu = 1$. It yields

$$\langle 1, 1 \rangle = 2$$

Reducing it with the relation (78) for which $s = 2, \nu = 0$ we get

$$\langle 2, 0 \rangle = 8 \quad (79)$$

The next truncated relation is $s = \nu = 2$, which yields

$$\langle 2, 2 \rangle = \frac{16}{3}$$

and the reduction gives

$$\begin{aligned} \langle 3, 1 \rangle &= \frac{92}{3} \\ \langle 1, 0 \rangle &= 184 \end{aligned} \quad (80)$$

Similarly in the next stage

$$\begin{aligned} \langle 3, 3 \rangle &= \frac{128}{7} \\ \langle 4, 2 \rangle &= \frac{18112}{133} \\ \langle 5, 1 \rangle &= \frac{421600}{399} \\ \langle 6, 0 \rangle &= \frac{3372800}{399} \end{aligned} \quad (81)$$

As an example of an average which cannot be had explicitly we may take the mean absolute value of the speed, that is $\langle 1, 0 \rangle$. We find this

value by picking a sequence of trial functions for $h_0(w)$ with the appropriate number of parameters and imposing successively (79), (80) and (81) upon this sequence; this leads us to a sequence of values for $\langle w \rangle$ which can then be examined. In such a procedure careful consideration of the trial functions is an important element. The following information is available. It was proved in Section IID that $h_0(w)$ is logarithmically infinite at the origin. At infinity, on the other hand, $h_0(w)$ falls as e^{-w} times some power of w . One way to check this is to drop the terms containing $1/c$ as factor in (46); the solution of the recursion system becomes then

$$h_r(w) \sim (2\nu + 1)e^{-w}w^k$$

where k is some unknown exponent. Armed with this fore-knowledge, we shall use the following sequence of trial function for $h_0(w)$

$$h_0(w) = pEi(w) + qK_0(w) + re^{-w} + swK_1(w) \quad (82)$$

where

$$Ei(w) = \int_w^\infty \frac{e^{-u}}{u} du$$

and $K_0(w)$, $K_1(w)$ are the modified Hankel functions of order zero and 1.²¹

We find in zeroth approximation from normalization only

$$\begin{aligned} p^{(0)} &= \frac{3}{2} & q^{(0)} &= r^{(0)} = s^{(0)} = 0 \\ \langle w \rangle^{(0)} &= 2.2500 \end{aligned} \quad (83a)$$

in first approximation, using (79)

$$\begin{aligned} p^{(1)} &= \frac{5}{6} \\ q^{(1)} &= \frac{4}{9} & r^{(1)} &= s^{(1)} = 0 \\ \langle w \rangle^{(1)} &= 2.3818 \end{aligned} \quad (83b)$$

²¹ This definition, which is in accord with the tables of Jahnke-Emde, differs from the usual one by a factor $2/\pi$. This change is suggested by Watson, Bessel Functions, p. 79, and proves convenient in the following.

in second approximation, using (79) and (80)

$$\begin{aligned} p^{(2)} &= \frac{7}{12} \\ q^{(2)} &= \frac{32}{45} \\ r^{(2)} &= -\frac{1}{20} \\ \langle w \rangle^{(2)} &= 2.3858 \end{aligned} \quad s^{(2)} = 0 \quad (83c)$$

and in third approximation, using (79), and (80) and (81)

$$\begin{aligned} p^{(3)} &= \frac{3079}{7980} \\ q^{(3)} &= \frac{202544}{209475} \\ r^{(3)} &= -\frac{2507}{18620} \\ s^{(3)} &= \frac{3152}{209475} \\ \langle w \rangle^3 &= 2.3864 \end{aligned} \quad (83d)$$

Appearances indicate strongly that the sequence (83) for $\langle w \rangle$ approaches a limit which one would guess to be

$$\langle w \rangle = 2.3865 \quad (84)$$

More evidence that the conclusion drawn is correct can be obtained by using the set of trial functions

$$h_0(w) = pK_0(w) + qe^{-w} + rwe^{-w}$$

We find then the following sequence of values for $\langle w \rangle$.

$$\langle w \rangle^{(0)} = 2.546 \quad \langle w \rangle^{(1)} = 2.395 \quad \langle w \rangle^{(2)} = 2.388$$

This descending sequence confirms (84) by approaching this same value from above.

Further evidence for the correctness of the procedure can be obtained by deriving a function $h_0(w)$ from the Monte Carlo function $h(\mathbf{w})$ discussed in Section IID and comparing it with our trial function. The function was constructed by covering Fig. 13 with a grid of concentric

circles and horizontal lines and replacing the integration

$$h_0(w) = 2\pi \int_{\vartheta=0}^{\vartheta=\pi} h(w) d(\cos \vartheta)$$

by a summation over grid points. The function $h_0^+(w)$ so obtained is compared in Table I with $h_0^{(0)}(w)$, $h_0^{(1)}(w)$ and $h_0^{(2)}(w)$ as defined by (82) and the numbers following. We observe that the first approximation

TABLE I

Comparison of the Monte Carlo $h_0^+(w)$ for $h_0(w)$ with successive approximations obtained by the new method.

w	$h_0^+(w)$	$h_0^{(0)}(w)$	$h_0^{(1)}(w)$	$h_0^{(2)}(w)$
0		∞	∞	∞
0.5	0.74	0.8397	0.7281	0.7144
1	0.29	0.3291	0.3019	0.3002
1.5	0.15	0.1500	0.1438	0.1440
2	0.081	0.0734	0.0730	0.0733
2.5	0.0412	0.0374	0.0384	0.0387
3	0.0199	0.0196	0.0207	0.0208
3.5	0.0118	0.0105	0.0114	0.0114
4	0.0063	0.0057	0.0063	0.0063
4.5	0.0030	0.0031	0.0035	0.0036
5	0.0014	0.0017	0.0020	0.0020
6	0.0004	0.0005	0.0007	0.0006
7	0.0001	0.0002	0.0002	0.0002

is an improvement over the zeroth one, while the second one makes little difference, considering the accuracy to which $h_0^+(w)$ is given. In individual cases the sequence drifts away from $h_0^+(w)$; this is not surprising because the latter function is very rough; this is to be expected from its mode of derivation.

The application of this method to the hard sphere model of ion-atom collisions offers no new feature of principle. The actual working out of results is somewhat more complicated, mainly because the connection diagram for the recursion system (49) is more involved. According to equation (26b) the dimensionless variable to be used in this work is

$$w = \frac{c}{\sqrt{a\lambda}} \quad (85)$$

We denote its averages $\langle w^s P, \langle \cos \vartheta \rangle \rangle$ by $\langle s, \nu \rangle$ as previously. The equa-

tion system (49) then takes the form

$$(2\nu + 1)(1 - \langle I_{s,\nu}(\chi) \rangle) \langle s + 1, \nu \rangle = \nu(\nu + s + 1) \langle s - 1, \nu - 1 \rangle + (\nu + 1)(s - \nu) \langle s - 1, \nu + 1 \rangle \quad (86)$$

The numbers $\langle I_{s,\nu}(\chi) \rangle$ were already discussed in connection with the system (78). What distinguishes (86) from (78) is the way in which the variables are connected; the new connection diagram which replaces Fig. 9 is shown in Fig. 14. The truncated relations no longer dovetail into each other as they did before. Only the first stage proceeds in a similar way, yielding explicit expressions for $\langle 2, 1 \rangle$ and $\langle 4, 0 \rangle$. In the next stage we

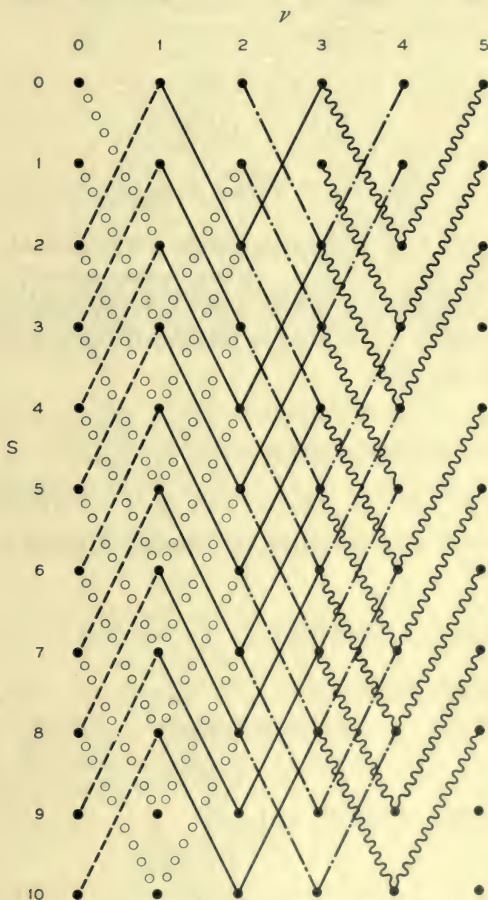


Fig. 14 — Interconnection established by the Boltzmann equation among the averages $\langle c^2 P_\nu (\cos \vartheta) \rangle$; case of constant mean free path.

start out with a relation between $\langle 1, 1 \rangle$ and $\langle 3, 2 \rangle$. By the use of regular recursion formulas we can successively transform this into a relation between $\langle 3, 0 \rangle$ and $\langle 3, 2 \rangle$, then $\langle 3, 0 \rangle$ and $\langle 5, 1 \rangle$ and finally between $\langle 3, 0 \rangle$ and $\langle 7, 0 \rangle$. Here we have for the first time the normal situation in which we do not get the actual value of a moment of $h_0(c)$ but only a relation between two or more of such moments; the reason for this is that the system fails to connect up with $\langle 0, 0 \rangle$ which equals unity a priori. A similar situation prevails for the next truncated relation; it is originally a relation between $\langle 2, 2 \rangle$ and $\langle 4, 3 \rangle$ and is finally reduced to one between $\langle 2, 0 \rangle$, $\langle 6, 0 \rangle$ and $\langle 10, 0 \rangle$. Similarly, the next truncated relation reduces to a relation between $\langle 5, 0 \rangle$, $\langle 9, 0 \rangle$ and $\langle 13, 0 \rangle$ and so forth. The first three of these reduced relations come out to be

$$\langle w^4 \rangle = 10 \quad (87)$$

$$3\langle w^7 \rangle = 112\langle w^3 \rangle \quad (88)$$

$$\frac{295}{56} \langle w^6 \rangle = 27\langle w^2 \rangle + \frac{17}{330} \langle w^{10} \rangle \quad (89)$$

These formulas will now be imposed upon a sequence of trial functions for $h_0(w)$ suitably chosen. Again, we may make use of the information of Section IID, according to which $h_0(w)$ is logarithmically singular at the origin. For large w we proceed as previously from (46) leaving off the terms of $1/c$. We get then

$$h_\nu(w) \sim (2\nu + 1)e^{-\frac{1}{2}w^2}w^k$$

This suggests the following trial function for $h_0(w)$

$$h_0(w) = pEi(\frac{1}{2}w^2) + qK_0(\frac{1}{2}w^2) + re^{-\frac{1}{2}w^2} + sw^2K_1(\frac{1}{2}w^2) \quad (90)$$

The best zero order approximation is actually obtained by the function $K_0(\frac{1}{2}w^2)$. We find

$$q^{(0)} = \frac{\pi}{2\Gamma^2(\frac{3}{4})} = 1.04605 \quad p^{(0)} = r^{(0)} = s^{(0)} = 0$$

In first order we get, using (87)

$$\begin{aligned} p^{(1)} &= -0.46543 \\ q^{(1)} &= 1.45285 \end{aligned} \quad r^{(1)} = s^{(1)} = 0$$

In the second order, using (87) and (88)

$$\begin{aligned} p^{(2)} &= -0.80856 \\ q^{(2)} &= 1.88127 \\ r^{(2)} &= -0.09804 \end{aligned} \quad s^{(2)} = 0$$

In third order, using (87) and (88) and (89)

$$p^{(3)} = -1.15071$$

$$q^{(3)} = 2.37034$$

$$s^{(3)} = 0.02062$$

$$r^{(3)} = -0.29016$$

These successive approximations lead to the following sequence for the drift velocity $\langle w \cos \vartheta \rangle$

$$\langle w \cos \vartheta \rangle^{(0)} = 1.04605 \quad (91a)$$

$$\langle w \cos \vartheta \rangle^{(1)} = 1.14256 \quad (91b)$$

$$\langle w \cos \vartheta \rangle^{(2)} = 1.14616 \quad (91c)$$

$$\langle w \cos \vartheta \rangle^{(3)} = 1.14661 \quad (91d)$$

We conclude from this sequence that

$$\langle w \cos \vartheta \rangle = 1.1467 \quad (92)$$

In addition to the drift velocity there is some interest in the energy and the energy partition. For the energy the following numbers are obtained

$$\langle w^2 \rangle^{(0)} = 2.1884 \quad (93a)$$

$$\langle w^2 \rangle^{(1)} = 2.3395 \quad (93b)$$

$$\langle w^2 \rangle^{(2)} = 2.3511 \quad (93c)$$

$$\langle w^2 \rangle^{(3)} = 2.3531 \quad (93d)$$

giving

$$\langle w^2 \rangle = 2.353 \quad (94)$$

A zero order value for $\langle w^2 \cos^2 \vartheta \rangle$ cannot be said to exist because the first truncated relation is the condition that a distribution function $h_2(w)$ exists at all. Thus, we can get only three numbers in a sequence approximating $\langle w^2 \cos^2 \vartheta \rangle$

$$\langle w^2 \cos^2 \vartheta \rangle^{(1)} = 1.8005 \quad (95a)$$

$$\langle w^2 \cos^2 \vartheta \rangle^{(2)} = 1.7696 \quad (95b)$$

$$\langle w^2 \cos^2 \vartheta \rangle^{(3)} = 1.7685 \quad (95c)$$

giving

$$\langle w^2 \cos^2 \vartheta \rangle = 1.768 \quad (96)$$

We can understand the results (92), (94) and (96) by giving the fraction of the total energy in ordered motion and the fraction of the energy in motion along the z -direction. We find for the first ratio

$$\frac{\langle w \cos \vartheta \rangle^2}{\langle w^2 \rangle} = 0.559 \quad (97)$$

and for the second

$$\frac{\langle w^2 \cos^2 \vartheta \rangle}{\langle w^2 \rangle} = 0.751 \quad (98)$$

The ratio (97) equals 0.5000 for all mean free time models; the ratio (98) is 0.778 for the mean free time model with isotropic scattering. Thus, the deviations from the earlier results are not drastic. However, in certain derived relations the difference is more noticeable. For instance, a good measure of the anisotropy of the diffusion process is furnished by the ratio of the random energy along the field to the energy at right angles.²² From (97) and (98) we find for this number

$$\frac{\langle w^2 \cos^2 \vartheta \rangle - \langle w \cos \vartheta \rangle^2}{\frac{1}{2}(\langle w^2 \rangle - \langle w^2 \cos^2 \vartheta \rangle)} = 1.54 \quad (99)$$

For the mean free time case this number equals 2.50. Hershey⁶ in his work assumes this number to be 1.000.

A comprehensive list of velocity averages is attached in Table II. As a comment I may add that the obvious mode of constructing such a table, namely by computing the column $\nu = 0$ from (90) and then using the recursion system (86) for the others, runs into some difficulty. First of all, a series of cancellations reduces the accuracy as ν increases; finally, at the positions marked "impossible" we find the missing third members of the truncated relations. These elements cannot be computed by recursion at all, but would require an explicit solution of the equation system (47) for $h_{\nu+1}(c)$. In the table, this more arduous path is not followed. Instead, the recursion method is used for the numbers in italic type and a few numbers are added by extrapolation. The numbers so obtained will be needed in Section IVB.

The calculations on the hard sphere model are immediately applicable to the experimental data of Figs. 3 to 7, which exhibit the drift velocity of

²² See below, equations (147) and (165).

the noble gas ions in the parent gas as functions of the parameter a/N . These data have a high field range in which the drift velocity varies as the square root of a/N . This is the variation for a model with constant mean free path, as seen from the dimensional formula (26b). It was indicated furthermore in the Section IA that we have good reason to think of the scattering between an ion and an atom as nearly isotropic.²³ These two features characterize uniquely the hard sphere model whose treatment we have just completed. To the extent that they are verified

TABLE II

Dimensionless high-field velocity averages $\langle s, v \rangle$ for the hard sphere model and mass ratio unity.

ν	0	1	2	3	4
s					
0	1.0000	0.7845	impossible		
1	1.3923	1.1467	0.8022	impossible	
2	2.3534	2.0000	1.4759	0.990	impossible
3	4.5868	3.9853	3.0578	2.134	
4	10.0000	8.8353	6.992	5.0602	3.474
5	23.912	21.405	17.330	12.84	
6	61.847	55.97	46.177	35.36	
7	171.241	156.3	130.91		
8	503.7	462.81			
9	1563	1445			
10	5090.9	4750			

the model is applicable to the experimental data. The formula to apply is (92) in combination with (85):

$$\langle c_z \rangle = 1.147 \sqrt{\frac{a}{N\sigma}} \quad (100)$$

In the logarithmic plot of $\langle c_z \rangle$ vs a/N the intercept of the straight line of slope 1/2 which fits the high field data thus equals

$$\log \frac{1.147}{\sqrt{\sigma}}.$$

The values for σ which result from this are shown in Table III. For comparison are shown the corresponding atomic cross section as determined from viscosity data.²⁴ It is interesting to observe that the ratio of

²³ A quantitative discussion of this point for the polarization force will follow in Section IIIB.

²⁴ Landolt-Börnstein, 1950 edition, Vol. I, part 1, page 325.

the two retains very nearly the constant value 3 throughout the table. The fact that the ratio is substantially larger than unity is explained by the resonance feature of the ion-atom scattering process as discussed in Section IA. The fact that it is constant is perhaps an indication of the fact that both processes are governed by overlap conditions of essentially the same wave functions.

I would like to point out in connection with the calculations of this section that the method developed is potentially of very wide application. One question that comes up, for instance, is whether a careful kinetics calculation is necessarily restricted to certain models or whether an ion-atom cross section known numerically could be used to derive therefrom kinetic properties. This is indeed possible. Suppose, for instance, that the cross section $\sigma(c)$ were available as a function of c for collision of He^+ -ions and He-atoms and suppose that this cross section were to satisfy the condition of isotropy $\Pi(\chi) = 1$ to a good approxi-

TABLE III

Cross sections for ion-atom and atom-atom collisions for the noble gases.

Gas	ion-atom cross section $\times 10^{16} \text{ cm}^2$	atom-atom cross section $\times 10^{16} \text{ cm}^2$
He	54	15.0
Ne	65	21.0
A	134	42.0
Kr	157	49
Xe	192	67

mation; we may then derive for this eventuality conditions on $h_0(c)$ which are more general, respectively, than (79) or (87), (80) or (88), (81) or (89). Since we are outrunning here the experimental evidence we shall limit ourselves to the derivation of the first of these relations. The first truncated relation is exactly (50a) which, for isotropy and equal masses, reads

$$\left\langle \frac{c_z}{a\tau(c)} \right\rangle = 2 \quad (101)$$

The reduction of this formula to a condition on $h_0(c)$ requires the relationship $\nu = 0$ of the set (46). This relation is always integrable to yield $h_1(c)$ in terms of $h_0(c)$, as was pointed out early in this section. For the special circumstances assumed the integrated equation is equation (74)

$$h_1(c) = 3 \int_c^\infty \frac{h_0(\gamma)}{a\tau(\gamma)} d\gamma \quad (102)$$

The elimination of $h_1(c)$ is achieved by forming the average (101) on the function (102). This leaves the required condition on $h_0(c)$; it may be given the following form

$$\left\langle \frac{\sigma(c)}{c} \int_0^c \gamma^4 \sigma(\gamma) d\gamma \right\rangle = 2 \frac{a^2}{N^2} \quad (103)$$

The equations (79) and (87) are manifestly special cases of this more general relation. Adaptations of this procedure to other cases are clearly possible whenever the need arises.

The calculations of this section are meant to suggest that it is possible to compute reliably average values from a Boltzmann equation without solving it completely. The method employed here for this purpose resembles a Ritz method in that it works with trial functions which must be guessed at, and like that method it is capable of indefinite improvement. The numerical results suggest strongly that we are converging toward a definite answer; however, a mathematical proof of this fact has not been presented. The method will be applied once more in the section on diffusion.

PART III — MOTION OF UNIFORM ION STREAMS IN INTERMEDIATE FIELDS

IIIA. A CONVOLUTION THEOREM

Whenever we deal with the motion of a given type of charged particle in a gas of given composition, then there exists a wide range of densities n and N as discussed in Section IA in which the motion of these particles depends only on a/N and kT . For this range the motion is governed by equation (13). Since deriving that equation, all our efforts were dealing with the "high field" equation (34) or (40), in which the gas temperature is taken to be zero and the electric field often scales out, as in (26), (75) and (85). The accomplished solution of this restricted problem, together with the low field solutions available in the literature, brings us back to the more general equation (13) and the question what can be done with it. The topic of Part III so defined is definitely inferior in importance to the one in Part II. For we are studying here an intermediate range of variables which can be handled qualitatively, both in concept and practice, by some sort of interpolation between the high and low field regimes. For precise measurements, conditions can always be chosen so as to satisfy one or the other of the two extremes. For this reason the intermediate field case will only be pushed as far as it will go conveniently, without appeal to numerical methods.

In this Section IIIA we shall give a complete solution of the inter-

mediate field problem for the mean free time models discussed in Section IIB. This solution is achieved by the following theorem: *Given the general equation (13) for constant mean free time*

$$a\tau \frac{\partial f(\mathbf{c})}{\partial c_z} + f(\mathbf{c}) = \frac{1}{4\pi} \iint M(\mathbf{C}') f(\mathbf{c}') \Pi(\chi) d\Omega_{\gamma'} d\mathbf{C} \quad (104)$$

and the "high field" equation derived from it by setting the gas temperature equal zero

$$a\tau \frac{\partial h(\mathbf{c})}{\partial c_z} + h(\mathbf{c}) = \frac{1}{4\pi} \iint \delta(\mathbf{C}') h(\mathbf{c}') \Pi(\chi) d\Omega_{\gamma'} d\mathbf{C} \quad (105)$$

and the Maxwellian equation derived from (104) by dropping the field term

$$m(\mathbf{c}) = \frac{1}{4\pi} \iint M(\mathbf{C}') m(\mathbf{c}') \Pi(\chi) d\Omega_{\gamma'} d\mathbf{C} \quad (106)$$

then the solution $f(\mathbf{c})$ of (104) is the convolution of the solution $h(\mathbf{c})$ of (105) and the solution $m(\mathbf{c})$ of (106):

$$f(\mathbf{c}) = \int h(\mathbf{u}) m(\mathbf{c} - \mathbf{u}) d\mathbf{u} \quad (107)$$

We carry through the proof by constructing explicitly the equation satisfied by the convolution. We replace the running variables $\mathbf{c}$, $\mathbf{c}'$, $\mathbf{C}$, $\mathbf{C}'$ in (105) by $\mathbf{u}$, $\mathbf{u}'$, $\mathbf{U}$, $\mathbf{U}'$ and multiply in $m(\mathbf{c} - \mathbf{u})$. We get

$$\begin{aligned} a\tau \frac{\partial h(\mathbf{u})}{\partial u_z} m(\mathbf{c} - \mathbf{u}) + h(\mathbf{u}) m(\mathbf{c} - \mathbf{u}) &= \\ &= \frac{1}{4\pi} \iint \delta(\mathbf{U}') h(\mathbf{u}') m(\mathbf{c} - \mathbf{u}) \Pi(\chi_u) d\Omega_{\eta'} d\mathbf{U} \end{aligned}$$

We now define $f(\mathbf{c})$ by the relation (107), and integrate the above equation over $\mathbf{u}$. The second member on the left comes out to be $f(\mathbf{c})$. For the first member, we carry out an integration by parts:

$$\begin{aligned} \int \frac{\partial h(\mathbf{u})}{\partial u_z} m(\mathbf{c} - \mathbf{u}) d\mathbf{u} &= - \int h(\mathbf{u}) \frac{\partial m(\mathbf{c} - \mathbf{u})}{\partial u_z} d\mathbf{u} \\ &= + \int h(\mathbf{u}) \frac{\partial (m(\mathbf{c} - \mathbf{u}))}{\partial c_z} d\mathbf{u} \\ &= \frac{\partial}{\partial c_z} \int h(\mathbf{u}) m(\mathbf{c} - \mathbf{u}) d\mathbf{u} \\ &= \frac{\partial f(\mathbf{c})}{\partial c_z} \end{aligned}$$

For the right hand member we observe that we have the eightfold integration

$$d\Omega_{\eta'} d\mathbf{U} d\mathbf{u},$$

that is an integration over the collision angles and all final velocity components. By a general principle of kinetic theory²⁵ we can invert in this integration the final and the initial quantities and write

$$d\Omega_{\eta'} d\mathbf{U} d\mathbf{u} = d\Omega_{\eta} d\mathbf{U}' d\mathbf{u}' \quad (108)$$

This puts us in a position to eliminate the δ -function by integration. We find

$$a\tau \frac{\partial f(\mathbf{c})}{\partial c_z} + f(\mathbf{c}) = \frac{1}{4\pi} \iint h(\mathbf{u}') m(\mathbf{c} - \mathbf{u}) \Pi(\chi_u) d\Omega_{\eta} d\mathbf{u}' \quad (109)$$

with the side condition that $\mathbf{u}$, $\mathbf{U}$, $\mathbf{u}'$, $\mathbf{U}'$ form a quadruple of vectors in the sense discussed in Section IB for which in addition

$$\mathbf{U}' = 0$$

If we substitute (107) into (104), denoting the dummy variable by $\mathbf{u}'$ instead of $\mathbf{u}$, then the two equations (104) and (109) take on a very similar look. A proof of their identity hinges upon proving the identity of the integral terms:

$$\begin{aligned} \int h(\mathbf{u}') d\mathbf{u}' \int m(\mathbf{c} - \mathbf{u}) \Pi(\chi_u) d\Omega_{\eta} \\ = \int h(\mathbf{u}') d\mathbf{u}' \iint M(\mathbf{C}') m(\mathbf{c}' - \mathbf{u}') \Pi(\chi_c) d\Omega_{\eta'} d\mathbf{C} \end{aligned} \quad (110)$$

The form of this relation suggests the assumption that the expressions are identical before integration over $\mathbf{u}'$; this assumption is proved by the events below. The complicated function $h(\mathbf{u}')$ thus disappears from the problem. The other such function, namely $\Pi(\chi)$ disappears then also; for it is by assumption arbitrary, hence could be replaced by a δ -function for a fixed, but arbitrary χ . The two sides of (110) must therefore be equal before we integrate over χ_u or χ_c , and the two χ 's are to be taken equal and fixed. Defining angles as shown in the spherical diagram Fig. 15 we thus get (110) in the form

$$\int m(\mathbf{c} - \mathbf{u}) d\epsilon = \iint M(\mathbf{C}') m(\mathbf{c}' - \mathbf{u}') d\phi d\mathbf{C} \quad (111a)$$

²⁵ See Reference 4, Section 3.52.

This is to be true with the side conditions

$$\mathbf{c} = \text{fixed} \quad (111b)$$

$$\mathbf{u}' = \text{fixed} \quad (111c)$$

$$\mathbf{U}' = 0 \quad (111d)$$

$$\chi_c = \chi_u = \chi = \text{fixed} \quad (111e)$$

Equation (111) is an identity involving only elementary functions. Thus the relation itself is in a sense elementary. Those who wish to believe it, may consider the theorem proved; for completeness, however, the proof of (111) will now follow.

Call the left side of (111a) X , the right side Y . To determine X , we substitute from (7) and (111d)

$$\mathbf{u} = \frac{m}{M+m} \mathbf{u}' + \frac{M}{M+m} \boldsymbol{\eta}$$

with

$$\boldsymbol{\eta}^2 = u'^2$$

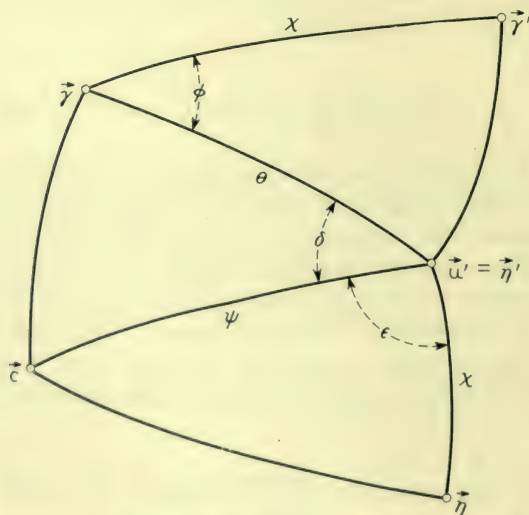


Fig. 15 — Definition of the angles occurring in the proof of the convolution theorem.

because of (9). This yields with the angles as shown on Fig. 15

$$X = \int_0^{2\pi} m(\mathbf{c} - \mathbf{u}) d\epsilon = \left(\frac{\beta m}{\pi}\right)^{3/2} \exp \left[-\beta m c^2 - \beta m u'^2 \frac{M^2 + m^2 + 2Mm \cos \chi}{(M + m)^2} + 2\beta m c u' \frac{m + M \cos \chi}{M + m} \cos \psi \right] \cdot \int_0^{2\pi} \exp \left[2\beta \frac{mM}{M + m} c u' \sin \chi \sin \psi \cos \epsilon \right] d\epsilon$$

The integral is evaluated by a formula known from the theory of Bessel functions

$$\int_0^{2\pi} e^{z \cos \epsilon} d\epsilon = 2\pi I_0(z) \quad (112)$$

and yields

$$X = \frac{2}{\sqrt{\pi}} (\beta m)^{3/2} \exp \left[-\beta m c^2 - \beta m u'^2 \frac{M^2 + m^2 + 2Mm \cos \chi}{(M + m)^2} + 2\beta m c u' \frac{m + M \cos \chi}{M + m} \cos \psi \right] \cdot I_0 \left(\frac{2\beta M m}{M + m} c u' \sin \chi \sin \psi \right) \quad (113)$$

Passing now to the right hand side of (111a) we may replace in the first place $d\mathbf{C}$ by $d\boldsymbol{\gamma}$, because of (111b). $\mathbf{c}'$ and $\mathbf{C}'$ are then replaced by the expressions

$$\mathbf{c}' = \mathbf{c} - \frac{M}{M + m} \boldsymbol{\gamma} + \frac{M}{M + m} \boldsymbol{\gamma}'$$

$$\mathbf{C}' = \mathbf{c} - \frac{M}{M + m} \boldsymbol{\gamma} - \frac{M}{M + m} \boldsymbol{\gamma}'$$

With the angles defined in Fig. 15, we thus get for Y

$$Y = \left(\frac{\beta M}{\pi}\right)^{3/2} \left(\frac{\beta m}{\pi}\right)^{3/2} \exp [-\beta M c^2 - \beta m(\mathbf{c} - \mathbf{u}')^2] \cdot \iiint \gamma^2 d\boldsymbol{\gamma} \sin \theta d\theta d\delta d\phi \cdot \exp \left[-\beta M \gamma^2 + \beta M c \gamma (\cos \psi \cos \theta + \sin \psi \sin \theta \cos \delta) - 2\beta \frac{mM}{M + m} u' \gamma (\cos \theta - \cos \theta \cos \chi - \sin \theta \sin \chi \cos \phi) \right]$$

Integrations over δ and ϕ again go with (112). Before writing down the result we shall pass over to a cylindrical coordinate system defined by

$$\begin{aligned}\gamma_{||} &= \gamma \cos \theta & \gamma_{\perp} &= \gamma \sin \theta \\ \gamma^2 d\gamma \sin \theta d\theta &= \gamma_{\perp} d\gamma_{\perp} d\gamma_{||}\end{aligned}$$

The result of the first two integrations then reads

$$\begin{aligned}Y &= \frac{4}{\pi} \beta^3 (Mm)^{3/2} \exp [-\beta M c^2 - \beta m(\mathbf{c} - \mathbf{u}')^2] \\ &\int_{-\infty}^{+\infty} d\gamma_{||} \exp \left[-\beta M \gamma_{||}^2 + 2\beta M \gamma_{||} \left(c \cos \psi - \frac{m u'}{M + m} (1 - \cos \chi) \right) \right] \\ &\int_0^{\infty} \gamma_{\perp} d\gamma_{\perp} \exp [\beta M \gamma_{\perp}^2] \cdot I_0(2\beta M c \gamma_{\perp} \sin \psi) \cdot I_0 \left(2\beta \frac{Mm}{M + m} u' \gamma_{\perp} \sin \chi \right)\end{aligned}$$

The first of these two integrals is elementary, the other is Weber's second exponential integral²⁶ which equals

$$\int_0^{\infty} \exp(-p^2 t^2) I_0(at) I_0(bt) t dt = \frac{1}{2p^2} \exp\left(\frac{a^2 + b^2}{4p^2}\right) I_0\left(\frac{ab}{2p^2}\right) \quad (114)$$

This yields for Y exactly the expression (113). The identity (111) is thus proved, and with it the convolution theorem.

The theorem just proved reduces the velocity distribution for arbitrary field and temperature to two components, one containing the field, but not the temperature, the other the temperature but not the field. In each of these components, in turn, the variable parameter scales out; thus the general distribution reduces to two basic ones one of which is the Maxwellian one: the other is worked out partially in the calculations of the Sections IIC and IID. The special case of heavy ion mass has been published independently by Kihara²⁷ without any apparent knowledge of this theorem which was available in the literature without complete proof.¹ Kihara's form of the theorem is that heavy ions in a light gas have an off-set Maxwellian distribution, with the gas temperature as parameter if the mean free time condition is obeyed for their collisions. Such a function is indeed the convolution of a Maxwellian distribution and the δ -function discussed in the Sections IIB and IIC.

The general distribution function resulting from (107) cannot be written down explicitly because this goal was never achieved for $h(\mathbf{c})$. However we do find a result which is almost a full substitute for this,

²⁶ Watson, G. N., *A Treatise on the Theory of Bessel Functions*. Cambridge University Press, Section 13.31, 1922.

²⁷ Kihara, Taro, *Rev. Mod. Phys.*, **24**, p. 45, 1952.

namely that all averages of products of integer powers of the Cartesian velocity components, which were shown to be computable in the high field case, can be computed for the intermediate and low field range as well. The calculation proceeds as follows. Suppose we wish to compute the velocity average

$$\langle c_x^m c_y^n c_z^p \rangle = \int c_x^m c_y^n c_z^p f(\mathbf{c}) d\mathbf{c} \quad (115)$$

for m, n, p integer or zero. We apply the convolution theorem (107) to $f(\mathbf{c})$, decompose the three factors into

$$c_x^m = \{u_x + (c_x - u_x)\}^m$$

$$c_y^n = \{u_y + (c_y - u_y)\}^n$$

$$c_z^p = \{u_z + (c_z - u_z)\}^p$$

and expand each of them by the binomial theorem. We find

$$\langle c_x^m c_y^n c_z^p \rangle = \sum_{\mu=0}^m \sum_{\nu=0}^n \sum_{\pi=0}^p \binom{\mu}{m} \binom{\nu}{n} \binom{\pi}{p} \quad (116)$$

$$\int h(\mathbf{u}) u_x^\mu u_y^\nu u_z^\pi d\mathbf{u} \int m(\mathbf{v}) v_x^{m-\mu} v_y^{n-\nu} v_z^{p-\pi} d\mathbf{v}$$

The second integral is a thermal average, the first a high field average computable by the method of Section IIB. Thus the average (116) is a finite sum of products of computable averages and is itself computable.

When formula (116) is applied to the averages (52), (53), (54), (57) and (59) very simple results are found because of the symmetry of the function $m(\mathbf{v})$. For the drift velocity $\langle c_z \rangle$ we get from (52)

$$\langle c_z \rangle = \frac{M+m}{M} \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle \quad (117)$$

This is the same formula as (52) which is thus proved to hold independently of the gas temperature. In the energy formulas we find simple addition of the thermal and high field values because the middle term in (116) drops out by symmetry. Inserting (53), (54), (57) and (59) we find

$$\langle mc^2 \rangle = 3kT + \frac{(M+m)^3}{M^2} \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle^2 \quad (118)$$

$$\langle mc_z^2 \rangle = kT + \frac{(M+m)^3 \left\langle \frac{M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle}{M^2 \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle} \quad (119)$$

$$\begin{aligned}
 \langle mc_z^2 \rangle - m\langle c_z \rangle^2 &= \\
 &= kT + \frac{(M + m)^2 \left\langle \frac{M \sin^2 \chi + 2m(1 - \cos \chi)^2}{a\tau} \right\rangle}{M \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle^2} \quad (120)
 \end{aligned}$$

$$\langle mc_x^2 \rangle = kT + \frac{(M + m)^3 \left\langle \frac{\sin^2 \chi}{a\tau} \right\rangle}{M \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{a\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{a\tau} \right\rangle^2} \quad (121)$$

The interpretation of these formulas is implicit in the discussion of the high field formulas given earlier. In particular the combination of the equations (55) and (116) can be given the elegant form

$$m\langle c^2 \rangle = M\langle C^2 \rangle + m\langle c_z \rangle^2 + M\langle c_x \rangle^2 \quad (122)$$

It states that the energy of an ion is obtained by adding the energy of a gas molecule, the energy visible in the drift motion and a storage term which is M/m times the energy in the drift motion; this term becomes important for electrons in a gas. A low field approximation to this formula (in which the second term on the right may be neglected) has been published in the article of Kihara.²⁸

IIIB. RESULTS FOR THE POLARIZATION FORCE AND THE ISOTROPIC "MAXWELLIAN" MODEL

The polarization force between ions and molecules which predominates over other forces at sufficiently low temperature satisfies the mean free time requirement of the preceding section. It follows that the complete theory given for those conditions applies to this force. The magnitude of force was given in (4). Its potential equals

$$V = \frac{1}{2} \frac{e^2 P}{\rho^4} \quad (123)$$

Classical theory is usually applicable to the scattering by the potential (123) because angular momentum quantum numbers run as high as 30 or 50 in normal situations.²⁹ This classical type theory, first developed by Langevin,³⁰ follows standard elementary methods for computing the

²⁸ Reference 27, formula 5.12.

²⁹ Holstein, Theodore, private communication, see also Reference 11.

³⁰ Langevin, Ann. de Chim. et de Phys., **5**, p. 245, 1905.

angle of deflection χ due to a potential of the type (123). The result is

$$\chi = \pi - 2 \int_0^{u_1} \frac{du}{\left\{ \frac{1}{b^2} - u^2 + \frac{e^2 P(M+m)}{Mmb^2\gamma^2} u^4 \right\}^{1/2}} \quad (124)$$

Here b is the "impact parameter", and u_1 is the lower of the two positive roots of the polynomial in the denominator; if the polynomial has no real root, the integration goes from 0 to ∞ . The question whether the denominator has a real root or not is tied up with the nature of the orbit. If b is sufficiently large a root exists and the orbit looks like a hyperbola, Fig. 2(a); for small b no root exists and the two particles are "sucked" towards each other in a spiralling orbit as shown in Fig. 2(b). The two regimes are separated by a limiting orbit in which the particles spiral asymptotically into a circular orbit. This limiting orbit is found by setting the discriminant of the square root in (124) equal to 0. We find

$$b_{\text{lim}}^4 = \frac{4e^2 P(M+m)}{Mm\gamma^2} \quad (125)$$

From this value of b_{lim} a cross section and a mean free time τ_s for spiralling collisions can be derived. We find

$$\tau_s = \frac{1}{2\pi eN} \left\{ \frac{Mm}{(M+m)P} \right\}^{1/2} \quad (126)$$

This is indeed a constant mean free time as stated, the speed of encounter γ having dropped out.

$1/\tau$ is the dimensional quantity entering into the averages $\langle \varphi(\chi)/\tau \rangle$ which occur in the Sections IIB and IIIA. In working them out in detail as was done by Hassé⁹ one has to take into account hyperbolic collisions also; for them a τ cannot be defined or comes out to be zero in the mean. This is due to small angle deflections which are infinitely probable. However, any quantity $\varphi(\chi)/\tau$ to be averaged in a physical problem contains a $\varphi(\chi)$ which vanishes for such impacts. Hence finite averages result which do not give overdue weight to these types of collisions. Following Hassé⁹, we do this in the following way for the present case. We write (124) in the form

$$\chi = \pi - 2 \int_0^{v_1} \frac{dv}{\left\{ 1 - v^2 + \frac{v^4}{4\beta^4} \right\}^{1/2}} \quad (127)$$

Here v equals bu and the parameter β equals b/b_{lim} . It is the parameter β introduced by Hassé. Now by the definition of τ we have

$$\begin{aligned}\left\langle \frac{\varphi(\chi)}{\tau} \right\rangle &= N \int \gamma \varphi(\chi) d\sigma(\gamma, \chi) \\ &= N \int_0^\infty \gamma \varphi(\chi) b db \int_0^{2\pi} d\epsilon \\ &= \pi N \gamma b_{\text{lim}}^2 \int_0^\infty \varphi(\chi) d(\beta^2)\end{aligned}$$

From (125) and (126) the factor in front of the integral just equals $1/\tau_s$; the integral on the other hand is a computable pure number independent of γ which is obtained by inserting into it the relationship (127) between χ and β . Hence we may write

$$\left\langle \frac{\varphi(\chi)}{\tau} \right\rangle = \frac{1}{\tau_s} \int_0^\infty \varphi(\chi) d(\beta^2) \quad (128)$$

The three equations (126), (127) and (128) completely define the nature of the averages appearing in previous sections. The integral (128) has to be computed by numerical methods. It is seen in the course of the evaluations that it naturally decomposes into two parts. The part for which β varies from 0 to 1 deals with spiralling collisions and exists for any $\varphi(\chi)$. For β between 1 and ∞ we get the contribution of the hyperbolic collisions to the average. This part is only finite if $\varphi(\chi)$ vanishes for small angle deflections.

The averages (52), (53), (54), (57) and (59), as well as (117) to (121) contain numerous averages of the form (128) all of which satisfy the predicted condition $\varphi(0) = 0$. They are obtained by linear combination of two basic types: $\langle (1 - \cos \chi)/\tau \rangle$ and $\langle \sin^2 \chi/\tau \rangle$. The first average is given in Hassé.⁹ Separating the parts due to spiralling and hyperbolic collisions we find

$$\begin{aligned}\int_0^1 (1 - \cos \chi) d(\beta^2) &= 0.8979 \\ \int_1^\infty (1 - \cos \chi) d(\beta^2) &= 0.2073\end{aligned}$$

This combines to give

$$\left\langle \frac{1 - \cos \chi}{\tau} \right\rangle = \frac{1}{\tau_s} \cdot 1.1052 \quad (129)$$

The analogous result for $\sin^2 \chi$ was obtained by the computing group of Bell Telephone Laboratories

$$\int_0^1 \sin^2 \chi d(\beta^2) = 0.511$$

$$\int_1^\infty \sin^2 \chi d(\beta^2) = 0.261$$

which gives

$$\left\langle \frac{\sin^2 \chi}{\tau} \right\rangle = \frac{1}{\tau_s} \cdot 0.772 \quad (130)$$

We may now rewrite the major results of Section IIIA for the polarization force. From equation (117) we get

$$\langle c_z \rangle = \frac{0.9048}{2\pi} \sqrt{\frac{1}{M} + \frac{1}{m}} \frac{E}{N\sqrt{P}} \quad (131)$$

This formula may be found in the literature.³¹ What is new about (131) is the realization that it is exact at high as well as low electric field.

The formula for the total energy needs no discussion for a special model; it does not involve the angular distribution law when written in the form (122). Thus we would obtain, for instance, for an ion travelling in the parent gas that its total energy is obtained by doubling its apparent energy observable in the drift and adding to this the thermal energy $\frac{3}{2}kT$.

For the partition of the high field component of the energy in the three coordinate directions we have two formulas, formula (58) partitions the entire field contribution of the kinetic energy, formula (60) only its random component. The first formula gives

$$e_x : e_y : e_z = M : M : (M + 6.73m) \quad (132)$$

Formula (60) gives

$$e_x : e_y : e_z^* = (M + m) : (M + m) : (M + 3.72m) \quad (133)$$

It is convenient to apply the general formulas also to the case of constant mean free time, coupled with the assumption of isotropic scattering. This combination of assumptions represents, strictly speaking, an im-

³¹ The formula is equation (3), p. 39 of Reference 2, in the limit $\lambda = 0$; or also the last unnumbered equation on p. 919 of Reference 6.

possibility; for we know of no mechanical force which realizes this arrangement. This model was already taken as the basis of the Monte Carlo calculation in Section IID. It will be seen now that it has a wider significance than one might anticipate. The necessary angular averages are

$$\langle 1 - \cos \chi \rangle = 1 \quad (134)$$

$$\langle \sin^2 \chi \rangle = \frac{2}{3} \quad (135)$$

This yields for (117)

$$\langle c_z \rangle = \frac{M + m}{M} a \tau \quad (136)$$

As usual, the formula for the energy does not involve the law of scattering if written in the form (122). If we choose the form (118) instead we get in agreement with (79)

$$\langle mc^2 \rangle = 3kT + \frac{(M+m)^3}{M^2} a^2 \tau^2 \quad (137)$$

The partition formula (58) becomes

$$e_x : e_y : e_z = M : M : (M + 6m) \quad (138)$$

the partition formula (60) which counts random energy only becomes

$$e_x : e_y : e_z^* = (M + m) : (M + m) : (M + 4m) \quad (139)$$

Comparison of these expressions with the ones for the polarization force shows that the difference between it and the isotropic model is remarkably small from a kinetic standpoint. We may see this by comparing (132) and (138) or (133) and (139). For the other formulas, we may compare more specifically the polarization results with an isotropic case having its mean free time τ given by

$$\tau = 0.9048 \tau_s \quad (140)$$

Equation (136) becomes then identical with (131) and because of (122) the same identity persists for the energy formula (137). In the light of this we may say that it is very nearly correct to state that scattering is isotropic for the polarization force. This qualitatively correct fact was repeatedly made use of in the preceding sections of the paper. The reason for it is chiefly the predominant effect of spiralling collisions. Indeed, equation (140) shows that a modification of τ_s by only 10 per cent takes into account the main influence of hyperbolic collisions.

From the discussion in Section IA it may be seen that the results

obtained for the polarization force have a potentially wide field of application when measurements of ion drift are extended to low temperature. In the meantime, the results apply occasionally at room temperature, whenever we deal with a small ion and are not bothered by special scattering mechanisms having large cross section. An example of this are the molecular noble gas ions in the parent gas whose drift velocities were measured by Hornbeck^{16, 17} and Varney.¹⁸ Table IV shows the measured mobility at standard gas density measured for these ions, in comparison with a value obtained from equation (131). The field range from which the observed mobility was obtained is intermediate. There is not only good numerical agreement, but the experiments follow the theory also in that there is little variation of the observed value

TABLE IV

Mobilities at standard density of the noble gas molecular ions. Comparison of the experiment with a formula based on the polarization force only.

Gas	$\mu_{\text{obs.}} \frac{\text{cm}^2}{\text{Volt sec}}$	$\mu_{\text{calc}} \frac{\text{cm}^2}{\text{Volt sec}}$
He	18	18.2
Ne	6.5	6.21
A	1.9	2.09
Kr	1.2	1.18
Xe	0.7	0.74

with the field. The discrepancy between the two columns can be used to determine a hard collision cross section which is to be superimposed on the polarization force, as is suggested in the so-called Langevin model.³⁰

IIIC. VELOCITY DISTRIBUTION FUNCTION FOR ELECTRONS

We have almost exhausted the results achieved for intermediate field conditions. For the sake of completeness I shall mention shortly the intermediate field distribution function for electrons whose derivation we owe to the ingenuity of Davydov.³²

The derivation does not differ in principle from the one presented in Section IIC for the electrons in the high field case. The distribution function is first expanded in spherical harmonics. For group theoretical reasons the scattering term in the Boltzmann equation is diagonal in

³² Davydov, B., Phys. Zeits. Sowjetunion., **8**, p. 59, 1935. See also Reference 4, pp. 349-350.

such a decomposition even in the presence of molecular agitation. Thence a generalized form of (47) may be derived containing essentially the same terms. Finally, all but the first two spherical harmonics are dropped and two equations analogous to (63) and (64) are obtained. In fact, it is found that equation (63) is maintained entirely. An extremely complicated reasoning is required, on the other hand, to find the generalization of (64). The result is

$$f_1(c) = \left\{ \frac{3kT}{M} \frac{df_0}{dc} + 3 \frac{m}{M} cf_0(c) \right\} \left\langle \frac{1 - \cos \chi}{a\tau(c)} \right\rangle \quad (141)$$

Combining (63) and (141) we find

$$\left(\left\langle \frac{1}{\frac{1 - \cos \chi}{a\tau(c)}} \right\rangle + \frac{3kT}{M} \right) \frac{df_0}{dc} + 3 \frac{m}{M} cf_0 = 0$$

and hence

$$f_0(c) = \exp \left[-m \int_0^c \frac{c \, dc}{\frac{\frac{1}{3} M}{\left\langle \frac{1 - \cos \chi}{a\tau(c)} \right\rangle^2} + kT} \right] \quad (142)$$

This is the so-called Davydov distribution which is a generalization containing within itself the Maxwellian distribution as well as the high field distribution (65).

The mean energy and the drift velocity of electrons may be calculated from (63) and (142). They are obtainable from the literature and will not be discussed here any further. Equipartition of the energy exists at all field conditions.

PART IV — DIFFUSIVE MOTION OF IONS

IVA. DIFFUSION FOR MEAN FREE TIME MODELS

It was proved in Section IC that if there are spatial inequalities in the distribution of the charge carriers then a smoothing out process sets in which can be described as diffusion. This derivation of principle can be supplemented for "Maxwellian" molecules by an explicit computation of the two components of the tensor (24), that is an evaluation of the integral (23). We shall do this by following the method of Maxwell¹⁹

rather than by generalizing the formal procedure of the Sections IIA, IIB and IIIA. Such a generalization would no doubt be possible, but would increase unduly the bulk of this paper. We shall operate therefore directly on equation (20). To get out the integral (23) we multiply the equation vectorially with $\mathbf{c}$ and integrate over $d\mathbf{c}$. This operation makes the first term vanish completely. This is obvious from symmetry for the components c_x and c_y of the multiplier $\mathbf{c}$. For c_z we have

$$a \int \frac{\partial g}{\partial c_z} c_z d\mathbf{c}$$

An integration by parts brings this in the form (18) and thus makes it equal to zero.

Temporarily, we may break the integral term of (20) into two parts, using some artificial procedure to eliminate small angle collisions. The first half of the integral term reads then simply

$$\frac{1}{\tau} \int g(\mathbf{c}) \mathbf{c} d\mathbf{c}$$

This is already the desired average (23). On the second half we use the identity (108) to give it the form

$$-\frac{1}{4\pi\tau} \iint M(\mathbf{C}') g(\mathbf{c}') \mathbf{c} \Pi(\chi) d\Omega_\gamma d\mathbf{C}' d\mathbf{c}'$$

We now use (7) to replace $\mathbf{c}$ by the expression

$$\mathbf{c} = \frac{m}{M+m} \mathbf{c}' + \frac{M}{M+m} \mathbf{C}' + \frac{M}{M+m} \boldsymbol{\gamma}$$

Only $\boldsymbol{\gamma}$ is affected by the integration over $d\Omega_\gamma$ which we take up first. Using $\boldsymbol{\gamma}'$ as the axis of a polar coordinate system we may write

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}_{||} + \boldsymbol{\gamma}_\perp$$

For every value of χ , $\boldsymbol{\gamma}_{||}$ has the fixed value $\boldsymbol{\gamma}' \cos \chi$. On the other hand the average of $\boldsymbol{\gamma}_\perp$ vanishes through integration over all azimuths. Thence we may write

$$\begin{aligned} \frac{1}{4\pi} \int \mathbf{c} \Pi(\chi) d\Omega_\gamma &= \frac{m}{M+m} \mathbf{c}' + \frac{M}{M+m} \mathbf{C}' + \frac{M}{M+m} (\mathbf{c}' - \mathbf{C}') \langle \cos \chi \rangle \\ &= \frac{m + M \langle \cos \chi \rangle}{M+m} \mathbf{c}' + \frac{M(1 - \langle \cos \chi \rangle)}{M+m} \mathbf{C}' \end{aligned}$$

We now multiply with $M(\mathbf{C}') g(\mathbf{c}')$ and integrate over $d\mathbf{c}' d\mathbf{C}'$. The integration of the term containing $\mathbf{C}'$ obviously vanishes for two independent reasons. The integration of the term in $\mathbf{c}'$, finally, yields again the average

(23). Combining the two pieces, we find

$$\frac{M}{M+m} \left\langle \frac{1 - \cos \chi}{\tau} \right\rangle \int g(\mathbf{c}) \mathbf{c} d\mathbf{c}$$

In this expression, the artificial exclusion of small angle scattering is no longer necessary and can be dropped. Completing the integrating of equation (20) we see that the right hand side gives averages over the unperturbed velocity distribution $f(\mathbf{c})$. Combining pieces, using (23) and indices 1, 2, 3 for the x , y and z components we get

$$\mathbf{j}_i(\mathbf{r}, t) = -n(\mathbf{r}, t) \sum_{v=1}^3 k_v \left[\frac{M+m}{M \left\langle \frac{1 - \cos \chi}{\tau} \right\rangle} \{ \langle c_i c_v \rangle - \langle c_i \rangle \langle c_v \rangle \} \right] \quad (143)$$

According to (16) and (24), the square bracket in (143) is the diffusion tensor. It has two distinct components which equal respectively

$$D_{||} = \frac{M+m}{M} \frac{\langle c_z^2 \rangle - \langle c_z \rangle^2}{\left\langle \frac{1 - \cos \chi}{\tau} \right\rangle} \quad (144)$$

$$D_{\perp} = \frac{M+m}{M} \frac{\langle c_x^2 \rangle}{\left\langle \frac{1 - \cos \chi}{\tau} \right\rangle} \quad (145)$$

The velocity averages entering are (120), and (121), that is the directional components of the random part of the energy. Substituting we get finally

$$D_{||} = \frac{(M+m)kT}{Mm \left\langle \frac{1 - \cos \chi}{\tau} \right\rangle} + a^2 \frac{(M+m)^3 \left\langle \frac{M \sin^2 \chi + 2m(1 - \cos \chi)^2}{\tau} \right\rangle}{M^2 m \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{\tau} \right\rangle^3} \quad (146)$$

$$D_{\perp} = \frac{(M+m)kT}{Mm \left\langle \frac{1 - \cos \chi}{\tau} \right\rangle} + a^2 \frac{(M+m)^4 \left\langle \frac{\sin^2 \chi}{\tau} \right\rangle}{M^2 m \left\langle \frac{3M \sin^2 \chi + 4m(1 - \cos \chi)}{\tau} \right\rangle \left\langle \frac{1 - \cos \chi}{\tau} \right\rangle^3} \quad (147)$$

The diffusion coefficients have the simple property that they are obtained by adding the low field and the high field limiting expressions.

This is a consequence of the limited form of the convolution theorem proved in Section IIIA; it probably implies also that the theorem can be extended in some form to include the case of diffusion.

It has been mentioned in the Section ID that the Nernst-Townsend relation (30) applies only to ions moving in a low field. We are now in a position to examine possible extensions of it to general fields. Equations (144), (145) and (117) suggest the form

$$\frac{D_n}{\text{mobility}} = \frac{2 \times \text{mean random energy along } n}{e} \quad (148)$$

where n stands for one of the principal directions of the diffusion tensor. This formula contains equation (30) as a specialization to the low field case.

Formula (148) is one of the formulas obtained in this study of ion motion in which model parameters do not appear. It is valid (a) for all interactions at low field and (b) for the mean free time case at all fields. It also holds dimensionally at high field for models obeying (25); this may be seen from (26a) and (28a). It appears a reasonable conjecture that (148) is approximately true for any law of interaction; the question will be taken up again in the next section.

Let us, in conclusion, write down the formulas resulting from (146) and (147) for the two special mean free time models studied in detail in Section IIIB: the polarization force and the isotropic model. The necessary averages are (129), (130), (134) and (135). They yield for the polarization force

$$D_{||} = \frac{M+m}{Mm} 0.905\tau_s \cdot kT + \frac{1}{3} \frac{(M+m)^3(M+3.72m)}{M^2m(M+1.908m)} a^2(0.905\tau_s)^3 \quad (149)$$

$$D_{\perp} = \frac{M+m}{Mm} 0.905\tau_s \cdot kT + \frac{1}{3} \frac{(M+m)^4}{M^2m(M+1.908m)} a^2(0.905\tau_s)^3 \quad (150)$$

and for the case of isotropic scattering

$$D_{||} = \frac{M+m}{Mm} \tau kT + \frac{1}{3} \frac{(M+m)^3(M+4m)}{M^2m(M+2m)} a^2\tau^3 \quad (151)$$

$$D_{\perp} = \frac{M+m}{Mm} \tau kT + \frac{1}{3} \frac{(M+m)^4}{M^2m(M+2m)} a^2\tau^3 \quad (152)$$

Just as in the earlier study the results for the two models do not differ appreciably.

IVB. LONGITUDINAL DIFFUSION FOR THE HARD SPHERE MODEL

Whenever the mean free time condition for collisions is not fulfilled, then the computation of diffusion coefficients requires a procedure analogous to that of Section IIE. Since this entails some numerical work the calculation was only carried out for a case which was thought to be of experimental interest, namely for longitudinal diffusion of ions in the parent gas. In other words, we are extending the numerical computation at the end of Section IIE to include longitudinal diffusion. The computation to provide us with the undetermined constant of equation (28b) for the special case when m and M are equal; it also offers, incidentally, a good test case for applying the method of Section IIE outside the area for which it was designed originally.

Since the equation is only to be solved in the high field case we may apply to (20) the reduction method of Section IIA. If we introduce also the specialization warranted by the hard sphere model and unit mass ratio then, in analogy to equation (40), we get the following starting equation

$$\frac{\partial g(\mathbf{w})}{\partial w_z} + wg(\mathbf{w}) - \frac{1}{4\pi} \int_0^\pi \frac{w'^4 \sin \chi d\chi}{w^3} \int_0^{2\pi} g(\mathbf{w}') d\omega \quad (153)$$

$$= -\lambda k \{w_z - \langle w_z \rangle\} h(\mathbf{w})$$

Here the dimensionless variable $\mathbf{w}$ defined by (85) has been employed instead of $\mathbf{c}$.

Equation (153) is the fundamental equation of our problem; it is an inhomogeneous version of equation (40). We solve the equation in the same way as we did previously, namely by decomposing $g(\mathbf{w})$ into spherical harmonics and forming moments. In other words we follow step by step the procedure of Section IIA, the only difference being the presence of an inhomogeneous term. We shall not enumerate all these steps again. We shall only note in passing the inhomogeneous form of (47) which is

$$\frac{1}{2} \int_0^\pi \frac{w'^4}{w^3} g_\nu(w') P_\nu(\cos \kappa) \sin \chi d\chi - wg_\nu(w)$$

$$- \frac{\nu}{2\nu - 1} \left\{ \frac{dg_{\nu-1}}{dw} - \frac{\nu - 1}{w} g_{\nu-1}(w) \right\}$$

$$- \frac{\nu + 1}{2\nu + 3} \left\{ \frac{dg_{\nu+1}}{dw} + \frac{\nu + 2}{w} g_{\nu+1}(w) \right\} =$$

$$= \lambda k \left[\frac{\nu}{2\nu - 1} w h_{\nu-1}(w) + \frac{\nu + 1}{2\nu + 3} w h_{\nu+1}(w) - \langle w_z \rangle h_\nu(w) \right]$$

Having introduced moments in the manner described earlier we arrive at the inhomogeneous version of (49) or (86)

$$\begin{aligned} \nu(s + \nu + 1)\{s - 1, \nu - 1\} + (\nu + 1)(s - \nu)\{s - 1, \nu + 1\} \\ - (2\nu + 1)(1 - \langle I_{s,\nu} \rangle)\{s + 1, \nu\} = - (2\nu + 1)\langle 1, 1 \rangle \langle s, \nu \rangle \quad (154) \\ + \nu \langle s + 1, \nu - 1 \rangle + (\nu + 1)\langle s + 1, \nu + 1 \rangle \end{aligned}$$

Here the curly brackets $\{s, \nu\}$ are normalized moments over $g(\mathbf{w})$ defined as follows

$$\{s, \nu\} = \frac{1}{k\lambda} \int g(\mathbf{w}) w^s P_\nu(\cos \vartheta) d\mathbf{w} \quad (155)$$

The equations (154) show that the quantities $\{s, \nu\}$ are numbers, the variable density gradient nk having been eliminated by the definition (155). The system does permit that arbitrary amounts of the pointed averages be added to the curly ones. This indeterminacy is removed by the supplementary condition (18) which, in the present notation reads

$$\{0, 0\} = 0 \quad (156a)$$

The connectivity of the equation system (154) is the same as that of (86). Hence it will have the same properties as that earlier system. We may, therefore, reduce it in the manner followed previously and get inhomogeneous versions of the equations (87), (88) and (89). They read

$$\{4, 0\} = -\frac{5}{3}\langle 4, 1 \rangle + \frac{5}{3}\langle 1, 1 \rangle \langle 3, 0 \rangle. \quad (157a)$$

$$\begin{aligned} & -\frac{10}{3}\langle 2, 0 \rangle - \frac{20}{3}\langle 2, 0 \rangle + 10\langle 1, 1 \rangle^2 \\ 112\{3, 0\} - 3\{7, 0\} & = 4\langle 7, 1 \rangle - 4\langle 1, 1 \rangle \langle 6, 0 \rangle \\ & + \frac{56}{5}\langle 5, 0 \rangle + \frac{112}{5}\langle 5, 2 \rangle \\ & - \frac{168}{5}\langle 1, 1 \rangle \langle 4, 1 \rangle \\ & - \frac{1344}{25}\langle 3, 1 \rangle + \frac{1344}{25}\langle 3, 3 \rangle \quad (158a) \\ & + \frac{448}{5}\langle 1, 1 \rangle \langle 2, 0 \rangle - \frac{448}{5}\langle 1, 1 \rangle \langle 2, 2 \rangle \end{aligned}$$

$$\begin{aligned}
54\{2, 0\} - \frac{295}{28}\{6, 0\} + \frac{17}{165}\{10, 0\} &= -\frac{17}{135}\langle 10, 1 \rangle + \frac{17}{135}\langle 1, 1 \rangle \langle 9, 0 \rangle \\
&- \frac{17}{36}\langle 8, 0 \rangle - \frac{17}{18}\langle 8, 2 \rangle + \frac{17}{12}\langle 1, 1 \rangle \langle 7, 1 \rangle \\
&+ 6\langle 6, 1 \rangle - \frac{21}{5}\langle 6, 3 \rangle - \frac{44}{5}\langle 1, 1 \rangle \langle 5, 0 \rangle + 7\langle 1, 1 \rangle \langle 5, 2 \rangle \quad (159a) \\
&+ \frac{54}{5}\langle 4, 0 \rangle + \frac{108}{7}\langle 4, 2 \rangle - \frac{288}{35}\langle 4, 4 \rangle \\
&- \frac{162}{5}\langle 1, 1 \rangle \langle 3, 1 \rangle + \frac{72}{5}\langle 1, 1 \rangle \langle 3, 3 \rangle
\end{aligned}$$

The pointed averages over the distribution $h(\mathbf{w})$ may be found in Table II. Substituting them we get

$$\{0, 0\} = 0 \quad (156b)$$

$$\{4, 0\} = -10.494 \quad (157b)$$

$$112\{3, 0\} - 3\{7, 0\} = 647.8 \quad (158b)$$

$$54\{2, 0\} - \frac{295}{28}\{6, 0\} + \frac{17}{165}\{10, 0\} = -566.4 \quad (159b)$$

The form (90) that was assumed for $h_0(w)$ will again be taken for $g_0(w)$ with new undetermined coefficients p, q, r, s and a factor $k\lambda$ evident from (153) or (155):

$$g_0(w) = k\lambda[pEi(\frac{1}{2}w^2) + qK_0(\frac{1}{2}w^2) + re^{-\frac{1}{2}w^2} + sw^2K_1(\frac{1}{2}w^2)] \quad (160)$$

This is a rather poor assumption because the form (90) was adopted for $h_0(w)$ after an extensive study of the properties of the distribution function $h(\mathbf{w})$. For $g(\mathbf{w})$ we know little beyond the fact that it is some kind of distorted p -type function. The $g_0(w)$ derived from this is not likely to resemble $h_0(w)$ very closely. Thus the choice (160) is mainly based on ignorance and convenience; this explains the slower convergence observed here than in (91), (93) and (95). To start with, the zero order is completely lost because (156) yields a zero coefficient. We find in first order, using (156) and (157)

$$\begin{aligned}
p^{(1)} &= 4.8842 & r^{(1)} &= s^{(1)} = 0 \\
q^{(1)} &= -4.2689
\end{aligned}$$

in second order, using (156), (157) and (158)

$$\begin{aligned} p^{(2)} &= -10.542 \\ q^{(2)} &= +14.993 & s^{(2)} &= 0 \\ r^{(2)} &= -4.408 \end{aligned}$$

in third order, using (156), (157), (158) and (159)

$$\begin{aligned} p^{(3)} &= -0.8710 \\ q^{(3)} &= +1.1754 \\ r^{(3)} &= +1.0140 \\ s^{(3)} &= -0.5809 \end{aligned}$$

The longitudinal diffusion coefficient results from these numbers by the use of (23), (24) and (160). With the notation (155) the formula becomes

$$D_{||} = -a^{1/2}\lambda^{3/2}\{1, 1\} \quad (161)$$

The formula (154) yielding $\{1, 1\}$ from $g_0(w)$ is $s = 2, \nu = 0$

$$\{1, 1\} = \frac{1}{4}\{3, 0\} + \frac{1}{2}\{3, 1\} - \frac{1}{2}\langle 1, 1 \rangle \langle 2, 0 \rangle \quad (162a)$$

or numerically from the Table II

$$\{1, 1\} = \frac{1}{4}\{3, 0\} + 0.6433 \quad (162b)$$

The result is

$$\{1, 1\}^{(1)} = -0.3695 \quad (163a)$$

$$\{1, 1\}^{(2)} = -0.2075 \quad (163b)$$

$$\{1, 1\}^{(3)} = -0.2198 \quad (163c)$$

The numbers do not extrapolate too reliably but one would guess that

$$\{1, 1\} = -0.22$$

is essentially correct. Hence we have

$$D_{||} = 0.22a^{1/2}\lambda^{3/2} \quad (164)$$

In order to gain an appreciation of the value obtained it is worthwhile to compare it with the value that would have been predicted from the generalized Nernst-Townsend relation (148). The mobility concept is ambiguous for all but the cases discussed then. It would seem that the appropriate concept here is the differential mobility because comparison

is made between a small density gradient and a small change in the applied field. Thus we would interpret (148) to mean

$$D_{||} \approx \frac{\partial \langle c_z \rangle}{\partial a} [\langle c_z^2 \rangle - \langle c_z \rangle^2] \quad (165a)$$

which, with (85), (92) and (96), becomes

$$D_{||} \approx 0.26a^{1/2}\lambda^{3/2} \quad (165b)$$

The error of formula (165) is thus 18 per cent, when compared to (164).

PART V — CONCLUDING OBSERVATIONS

The present article is supposed to contain the essentials of a kinetic theory of charged particles moving through a gas in the presence of an intermediate or high electric field. An effort was made to make the theory general, yet many irksome restrictions will become apparent to those who will try to apply it to their particular problem. Especially those who have in mind application to electrons will find the article unsatisfactory. It is true that many sections leave the masses variable; however, the assumption of elastic collisions, which is made throughout, is almost fatal to all but the most elementary applications. Thus most of the material is slanted for ions. Within this domain, numerous awkward restrictions are still found here. The most important ones are presumably the restriction to D.C. conditions, the assumption of "low" ion density, and the omission of all magnetic effects. It is my general impression, which I gained from the convolution theorem Section IIIA and which is confirmed by a recent publication²⁷ that much can be done to remove these three restrictions provided the mean free time assumption is made for collisions. To many the adoption of the mean free time condition will in itself appear an awkward restriction. In a rigorous sense this is true, and calculations are made in this article for the more appropriate hard sphere model when quantitative comparison with experiment is contemplated (equations (100) and (164)). Indications are even given for a treatment which dispenses altogether with the use of models (equation (103)). However, for rapid advance and easy handling, the mean free time assumption does appear essential. It is therefore important to point out that in a wider semiquantitative sense, the use of this model is no barrier to application. In other words, there is in the mean free time formulas information which suggests a wider validity. This is particularly true for equations which do not contain model parameters, such as (55), (56), (122) and (148). Even formulas which

do contain the mean free time yield to judicious treatment. For example, we have the hard sphere formula (100) for the drift velocity of an ion. This formula happens to be limited to the high field case and mass ratio unity. On the other hand we have formula (136) which holds for all fields and all mass ratios, but assumes constant mean free time. We now adopt this formula as a general guess for the hard sphere model, interpreting τ as previously as the mean free time between collisions; this quantity is now no longer a constant, but should be taken as

$$\tau = \frac{\lambda}{\sqrt{\langle c^2 \rangle + \langle C^2 \rangle}} \quad (166)$$

The denominator is the root mean square relative velocity which is familiar from other applications. The interpretation (166) yields a tentative formula for the drift for all mass ratios and for all fields. Specializing to the high field case, we may neglect $\langle C^2 \rangle$ in (166) and then substitute for $\langle c^2 \rangle$ from (55). This yields the high field formula

$$\langle c_z \rangle \approx \frac{(M + m)^{1/4} m^{1/4}}{M^{1/2}} (a\lambda)^{1/2} \quad (167)$$

This is indeed a very successful formula. For ions in the parent gas it differs from (100) by only 4 per cent. For electrons it checks the result of Druyvesteyn⁴ to within 12 per cent. Finally, for heavy ions in a light gas, we find exact agreement with equation (71). As a second specialization we may apply (166) to the low field case. We must then set

$$\langle c^2 \rangle + \langle C^2 \rangle = 3kT \left(\frac{1}{m} + \frac{1}{M} \right)$$

and get from (136) and (166)

$$\langle c_z \rangle = \frac{1}{\sqrt{3}} \left(\frac{1}{m} + \frac{1}{M} \right)^{1/2} \frac{eE\lambda}{\sqrt{kT}} \quad (168)$$

All dimensional factors in this formula are correct. Numerically (168) is somewhat inferior to (167); for the factor differs from the correct one³³ by 20 per cent. Nevertheless, the combination of (136) and (166) gives results which are semiquantitatively correct in all relevant limiting cases. This makes it a reliable interpolation formula for intermediate field conditions; for this case $\langle c^2 \rangle$ would have to be substituted from (122) and the resultant quadratic equation solved for $\langle c_z \rangle$.

From the examples given we may conclude that the mean free time formulas contain in essence information applicable to other types of elastic scattering.

³³ See Reference 2, page 40, second equation.

PART VI — ACKNOWLEDGEMENTS

This article is the outcome of years of fruitful cooperation with the gas discharge group of Bell Telephone Laboratories, and Dr. J. A. Hornbeck in particular. At one stage of the work I enjoyed the stimulation of Dr. R. W. Hamming who is responsible for all the details of the Monte Carlo calculation. Further acknowledgements are due to Miss C. L. Froelich who carried out the computation of the number in (130), and Miss M. Murray who carried a good share of the burden in the preparation of the manuscript. Finally, I express my thanks to Dr. K. G. McKay for his critical perusal of the manuscript.

Abstracts of Bell System Technical Papers* Not Published in This Journal

Experimental Verification of the Theory of Laminated Conductors. H. S. BLACK¹, C. O. MALLINCKRODT⁵ and S. P. MORGAN¹. *I.R.E., Proc.*, **40**, pp. 902-905, August, 1952.

Clogston has discovered that if a conductor is properly laminated, there exists a particular phase velocity along the conductor for maximum penetration of the fields and minimum loss due to skin effect. An experimental coaxial line was constructed whose center conductor was laminated and whose phase velocity could be varied by changing the dielectric constant of the main dielectric. As predicted by theory, the measured attenuation was critically dependent upon phase velocity. With optimum phase velocity the attenuation, though greater than predicted by theory, was less than that of a conventional coaxial cable of the same dimensions and same main dielectric. A theoretical analysis of the experimental laminated conductor is described in an appendix.

ASTM Standards—Their Effect on Plastics Technology. R. BURNS¹. *A.S.T.M. Bull.*, No. 183, pp. 78-80, July, 1952.

Typical Block Diagrams for a Transistor Digital Computer. J. H. FELKER¹. *A.I.E.E., Trans., Commun. & Electronics Sect.*, No. 1, pp. 175-182, July, 1952.

The first electric digital computers were built around the properties of relays. The superior speed capabilities of vacuum tubes has led in recent years to their use in new computer designs to replace relays. Because of the small size, low power consumption, and expected long life of transistors, it now appears that the transistor will replace the vacuum tube as a computer element. This paper presents a study of binary computer functions with recommended mechanizations that were selected because they appeared to be readily attainable with transistors now under development. Block diagrams are presented of switches, memory units, arithmetic units, and other basic components. Estimates are given for the number of parts required in the units. It is concluded that a high-performance all-semiconductor computer can be built with germanium diodes and transistors.

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

¹ Bell Telephone Laboratories.

⁵ Hughes Aircraft Company, Culver City, California

A Broad-Band Interdigital Circuit for Use in Traveling-Wave-Type Amplifiers. R. C. FLETCHER¹. *I.R.E., Proc.*, **40**, pp. 951-958, August, 1952.

Because of its high power-handling capacity, the interdigital circuit has been considered for use in traveling-wave-type amplifiers. An analysis is presented here which indicates that this type of circuit can be arranged to give constant phase velocity over a wide bandwidth (30 per cent), as required to give constant gain. The analysis is qualitatively checked experimentally. The impedance parameter (proportional to the cube of the gain in db) is approximately the same as the flattened helix (such as has been used for the magnetron amplifier) and about one-third that of the conventional circular helix.

Copper as an Acceptor Element in Germanium. C. S. FULLER¹ and J. D. STRUTHERS¹. *Phys. Rev.*, **87**, pp. 526-527, August 1, 1952.

Properties of Thermally Produced Acceptors in Germanium. C. S. FULLER¹, H. C. THEUERER¹ and W. VAN ROOSBROECK¹. *Phys. Rev.*, **85**, pp. 678-679, February 15, 1952.

Initial Permeability and Related Losses in Ferrites. J. K. GALT¹. *Ceramic Age*, **60**, pp. 29-33, August, 1952.

Three-Phase Power From Single-Phase Source. A. L. HOLCOMB⁶, *S.M.P.T.E., Jl.*, **59**, pp. 32-39, July, 1952.

Described is the development of a nonrotating device for the conversion of single-phase 115-volt power to a three-phase 230-volt form for the synchronous operation of cameras, sound recorders and other film pulling mechanisms associated with production of motion pictures.

Dominant Wave Transmission Characteristics of a Multimode Round Waveguide. A. P. KING¹. *I.R.E., Proc.*, **40**, pp. 966-969, August, 1952.

This paper presents some dominant wave transmission characteristics of multimode round waveguide lines in the 4-kmc range of frequencies. The use of such waveguide lines offers the advantages of lower transmission losses than obtainable with single-mode rectangular waveguide, and relative ease of making good joints. Possible mode conversion effects, including dominant mode elliptical polarization, have been examined and found to be innocuous. As a result, cross-polarized dominant waves can be used to provide two reasonably independent signaling channels at the same frequency in one pipe. The experimental results obtained with a straight line 2.812-inch inside diameter and length of 150 feet are given.

Superconductivity in the Cobalt-Silicon System. B. T. MATTHIAS¹. Letter to the Editor. *Phys. Rev.*, **87**, p. 380, July 15, 1952.

¹ Bell Telephone Laboratories

⁶ Westrex

Simple Phase-Angle Measurement Technique. J. A. RUDISILL, JR.³. *Electronics*, **25**, pp. 228, 232, 236, September, 1952.

Solid State Physics in Electronics and in Metallurgy. W. SHOCKLEY¹. *Jl. Metals*, **4**, pp. 829-842, August, 1952. (Monograph 2011).

Impurity Effects in the Thermal Conversion of Germanium. W. P. SLICHTER¹ and E. D. KOLB¹. *Phys. Rev.*, **87**, pp. 527-528, August 1, 1952.

Traffic Engineering Design of Dial Telephone Exchanges. J. A. STEWART⁴. *Midwest Engr.*, **4**, pp. 3-5, 17-18, May, 1952.

Single-Crystal Germanium. G. K. TEAL¹, M. SPARKS¹ and E. BUEHLER.¹ *I.R.E., Proc.*, **40**, pp. 906-909, August 1952.

Significant advances have been made in the development of new types of transistors, photocells, and rectifiers and in the improvement of the reproducibility and reliability of the point-contact transistor. A key factor in this development has been the use of single-crystal germanium having a high degree of lattice perfection and compositional control. Of particular interest to the device-development engineer is the fact that the rectifying barriers between the *p*-type and *n*-type sections behave in a manner predictable from the measured properties of each section. The exceptionally long lifetime of injected carriers observed in the material and the high degree of control over its chemical composition make it ideally suitable for the production of *p-n* structures. The ranges of properties of germanium single crystals which are now realizable are given, as well as their present degree of control.

Lead-Acid Stationary Batteries. U. B. THOMAS¹. *Electrochem. Soc. Jl.*, **99**, pp. 238C-241C, September, 1952.

Polymorphism of $ND_4D_2PO_4$. E. A. WOOD¹, W. J. MERZ¹ and B. T. MATTHIAS¹. *Phys. Rev.*, **87**, p. 544, August 1, 1952.

Lightning Protection for Mobile Radio Fixed Stations. D. W. BODLE¹. *I.R.E. Trans., P.G.V.C.-1*, pp. 122-133, February, 1952.

Equipment in fixed stations of a mobile radio system is susceptible to damage from lightning strokes to either the antennas or the connecting power and land communication facilities unless special protection is provided. The problem, however, is not alone one of protecting the station equipment, but consideration must also be given to the protection of these connecting facilities to insure their continuity of service. The causes of and factors affecting lightning damage are discussed, including the probable incidence of strokes to the antennas. General protection principles are outlined, and the application of specific protection methods is described.

¹ Bell Telephone Laboratories

³ Western Electric Company

⁴ Illinois Bell Telephone Company

Matching Coax Line to the Ground-Plane Antenna. R. T. DeCAMP². *QST*, **36**, pp. 18-19, 120, 122, September, 1952.

This article describes a method for predetermining antenna and matching-stub dimensions for matching any selected transmission line. Although applied particularly to the ground-plane antenna, the curves are useful for half-wave dipoles if allowance is made, when necessary, for the effect of ground on the antenna characteristics.

The Telephone System in National Defense. C. M. MAPES². *Military Engr.*, **44**, pp. 375-377, September-October, 1952.

Multi-Element Directional Couplers. S. E. MILLER¹ and W. W. MUMFORD¹. *I.R.E., Prcc.*, **40**, pp. 1071-1078, September 1952.

It is shown that the backward wave in a directional coupler is related to the shape of the function describing the coupling between transmission lines by the Fourier transform. This facilitates the design of directional couplers for arbitrary directivities over any prescribed frequency band. Tightly coupled directional couplers are analyzed in simple terms, and it is shown that any desired loss ratio, including complete power transfer between lines, may be achieved. The theories are verified using waveguide models operating at 4,000, 24,000, and 48,000 mc, and it is indicated that the work is applicable to many types of electrical and acoustic transmission lines.

Segregation of Two Solutes, with Particular Reference to Semiconductors. W. G. PFANN¹. *Jl. Metals*, **4**, pp. 861-865, August, 1952. (Monograph 2020).

The simultaneous segregation of two solutes during the directional solidification of an ingot is treated mathematically on the basis of simplifying assumptions. Expressions are derived for the difference in concentration of two solutes, and for the location and concentration gradient of a pn barrier formed in a semiconductor by the segregation of a donor and an acceptor.

Nonsynchronous Time Division with Holding and with Random Sampling. J. R. PIERCE¹ and A. L. HOPPER¹. *I.R.E., Proc.*, **40**, pp. 1079-1088, September, 1952.

There is a general type of system in which an indefinitely large number of transmitters can have access to any of an indefinitely large number of receivers over a medium of limited band-width. In these systems, signal-to-noise ratio goes down as more transmitters are used simultaneously. This paper describes a particular system which sends samples by means of coded pulse groups sent at random times. The signal-to-noise ratio is good in the absence of interference and the effect of interference is minimized by holding the previous sample if a sample is lost. An experimental system worked satisfactorily and gave close to the predicted signal-to-noise ratio might be used to provide communication and automatic switching in rural telephony, or for other applications.

¹ Bell Telephone Laboratories

² American Telephone and Telegraph Company

An Improved Electrolysis Switch. V. B. PIKE¹. *Corrosion*, **8**, pp. 311-313; disc. pp. 322-323, September, 1952. (Monograph 2021).

An improved electrolysis switch has been placed in use in the Bell System for mitigation of electrolysis of cables by stray currents. It consists of three relays operating in sequence, namely control, intermediate and drain relays. It automatically closes a drainage bond between a lead-covered underground cable and a power return ground when stray current is picked up by the cable to such an extent as to make some of the sheath positive with respect to its environment. It also opens bond when the drainage current falls to zero and drainage is no longer required. Two sizes are used capable of draining 200 amperes and 400 amperes, respectively. Its action is very fast, only 0.015 second elapsing from the time the control circuit releases until opening of the drainage bond. Separate adjustments are available for setting the voltage at which the switch closes and opens the drainage bond. A capacitive voltage booster enables switch operation over longer battery power supply wires than is possible if the booster is not used. A power supply unit consisting of a stepdown transformer and selenium rectifier is also available to operate the switch with power drawn from an AC power source if the switch must be installed beyond reach of the battery power. A sealed steel housing enables the switch to withstand submersion and permits installation in very damp locations.

Statistics of the Recombinations of Holes and Electrons. W. SHOCKLEY¹ and W. T. READ, JR.¹. *Phys. Rev.*, **87**, pp. 835-842, September 1, 1952. (Monograph 2022).

The statistics of the recombination of holes and electrons in semiconductors is analyzed on the basis of a model in which the recombination occurs through the mechanism of trapping. A trap is assumed to have an energy level in the energy gap so that its charge may have either of two values differing by one electronic charge. The dependence of lifetime of injected carriers upon initial conductivity and upon injected carrier density is discussed.

Motion of Gaseous Ions in a Strong Electric Field. G. H. WANNIER¹. *Phys. Rev.*, **87**, pp. 795-798, September 1, 1952. (Monograph 2023).

This paper continues an earlier one on the same subject. Its object is to elucidate the nature of the random motion of an ion about its drift. In Section F it is shown that this motion can be described as a diffusion with a diffusion tensor axially symmetric about the field. If the mean free time between the collisions of an ion with molecules is independent of speed, then explicit expressions may be derived for the two diffusion coefficients; these expressions are written down without proof in Section G; they are connected with the mobility by a natural extension of the Einstein relation. In Section H, the longitudinal diffusion coefficient is computed numerically for the hard sphere model, high field, and mass ratio 1; the method of computation is the same as in Section D. Finally, it is shown in Section I how approximate formulas of wider validity can be inferred from the ones obtained.

¹ Bell Telephone Laboratories

High-Frequency Crystal Units for Primary Frequency Standards. A. W. WARNER¹. *I.R.E., Proc.*, **40**, pp. 1030-1033, September, 1952.

A new approach to the design of crystal units for primary frequency standard use has resulted in crystal units in the 3- to 20-mc frequency range characterized by high Q and low capacitance in the series arm of the equivalent electrical circuit. By utilizing the overtone frequency of specially shaped AT-cut quartz plates, both Q and the rate of impedance change with frequency are enhanced together, and in addition the stability with time of the crystal unit is increased because of a larger frequency-determining dimension. Additional characteristics of the crystal units include small size, stability under conditions of vibration and shock, and low-temperature coefficient. Crystal-oscillator stabilities of one part in 10^8 per month have been achieved without recourse to stabilized circuits.

¹ Bell Telephone Laboratories

Contributors to this Issue

WALLACE C. BABCOCK, A.B., Harvard University, 1919; S.B., Harvard University, 1922. U. S. Army, 1917-19. American Telephone and Telegraph Company, 1922-34; Bell Telephone Laboratories, 1934-. Mr. Babcock was engaged in crosstalk studies until World War II, when he studied radio countermeasure problems for the N.D.R.C. Since then he has been concerned with antenna development for mobile radio and point-to-point radio telephone systems and has been engaged in other systems studies. Member of I.R.E. and Harvard Engineering Society.

JOHN BARDEEN, B.S., in E.E., University of Wisconsin, 1928; M.S. in E.E., University of Wisconsin, 1929; Ph.D., Princeton University, 1936. Gulf Research and Development Company, 1930-33; Harvard University, 1935-38; University of Minnesota, 1938-41; Naval Ordnance Laboratory, 1941-45; Bell Telephone Laboratories, 1945-51; University of Illinois, 1951-. At Bell Telephone Laboratories, Dr. Bardeen, co-inventor with Dr. Walter Brattain of the point-contact transistor, was primarily concerned with theoretical problems in solid state physics, including the study of semi-conductors, diffusion in solids, and superconductivity. Associate editor of *The Physical Review*, 1949-51. Stuart Ballantine Medal of the Franklin Institute, 1952. Fellow, American Physical Society; member, American Association for the Advancement of Science.

WALTER H. BRATTAIN, B.S., Whitman College, 1924; M.A., University of Oregon, 1926; Ph.D., University of Minnesota, 1929. Bureau of Standards, 1928-29; Bell Telephone Laboratories, 1929-. Dr. Brattain, co-inventor with Dr. John Bardeen of the point-contact transistor, has been primarily concerned with the study of semi-conductors at Bell Laboratories. During World War II he worked for the Division of War Research of Columbia University and is currently spending the fall term of the academic year 1952-53 as a visiting lecturer at Harvard University. Stuart Ballantine Medal of the Franklin Institute, 1952. Fellow, American Physical Society and American Association for the Advancement of Science; member, Sigma Xi and Phi Beta Kappa.

KENNETH BULLINGTON, B.S. in E.E., University of New Mexico, 1936; S.M., Massachusetts Institute of Technology, 1937; Bell Telephone Laboratories, 1937-. Until World War II, Mr. Bullington was occupied with systems engineering work on wire transmission circuits. Since 1942, he has been concerned with transmission engineering on radio systems, especially with radio propagation studies. Member of I.R.E., Phi Kappa Phi, Sigma Tau, and Kappa Mu Epsilon.

R. H. COLLEY, A.B., Dartmouth College, 1909; A.M., Harvard University, 1912; Ph.D., George Washington University, 1918; Austin Teaching Fellow in Botany, Harvard University, 1910-12; Instructor in Botany, Dartmouth College, 1909-10 and 1912-16; Pathologist, Division of Forest Pathology, Bureau of Plant Industry, U. S. Department of Agriculture, 1916-28. Bell Telephone Laboratories, 1928-1952. Dr. Colley was chairman of Committee 05—Wood Poles, of the American Standards Association for nearly twenty years. He was president of the American Wood-Preservers' Association 1943-44. During his years with the Laboratories he worked particularly on development and research problems connected with material and preservative treatment specifications for poles and other timber products used in outside plant. His more recent activities were directed toward improvement of laboratory techniques for evaluating wood preservatives, and toward the development of a coordinated plan for fundamental research on oil preservatives. He was Timber Products Engineer for the Laboratories from 1940 to 1950, and Timber Products Consultant from 1950 to 1952. His article in this issue of the JOURNAL was prepared before his retirement on May 31, 1952.

KARL K. DARROW, B.S., University of Chicago, 1911. He studied at the Universities of Paris and Berlin in 1911 and 1912, specializing in physics and mathematics; Ph.D., University of Chicago, 1917. He then joined the staff of Bell Telephone Laboratories, at that time known as the Engineering Department of Western Electric Company. Here his work has included the study, correlation, and representation of scientific information for his colleagues, keeping them informed of current advances made by workers in fields related to their own activities. As a corollary to his work, Dr. Darrow appears from time to time before scientific and lay audiences to lecture on current topics in physics and the related sciences. He has taken an active interest in education, teaching physics during summer and other sessions at Stanford, Chicago, and Columbia Universities and at Smith College. From 1944 to 1946, he served

as consultant to the Metallurgical Laboratory in Chicago. Dr. Darrow is the author of *Introduction to Contemporary Physics* (1926 and 1939), *Electrical Phenomena in Gases* (1932), *Resonance of Physics* (1936), and *Atomic Energy* (1948) and of many articles in this and other journals. He is a member of the American Physical Society, which he has served as secretary since 1941, the Physical Society of London, Société Française de Physique, the American Philosophical Society, of which he was a counsellor for four years. From 1949 to 1951 he was vice-president of the International Union of Pure and applied Physics. In 1949 he received an honorary doctorate from the Université de Lyon and was made Chevalier de la Légion d'Honneur in 1951.

JOHN RIORDAN, B.S., Sheffield Scientific School of Yale University, 1923. United Electric Light and Power Company, now a part of the Consolidated Edison Company, 1923-1926. Department of Development and Research of the American Telephone and Telegraph Company, 1926-1934. Bell Telephone Laboratories, 1934-. With the American Telephone and Telegraph Company, his work was largely on circuit and transmission theory, particularly in relation to inductive interference from electrified railways. This work was continued in Bell Telephone Laboratories after the Development and Research Department was consolidated with it in 1934. Since 1940 Mr. Riordan has been engaged in mathematical work: Boolean algebra in switching, number theory in cable splicing, and combinatorial and probability studies of traffic. Mr. Riordan is a member of the American Mathematical Society, the Mathematical Association of America, the Institute of Mathematical Statistics, and is a Fellow of the American Association for the Advancement of Science.

GREGORY H. WANNIER, Louvain University, 1930-31; University of Cambridge, 1933-34; Ph.D., University of Basel, 1935. Assistant, University of Geneva, 1935-36; Swiss-American Exchange Fellow, Princeton University, 1936-37; instructor, University of Pittsburgh, 1937-38; assistant lecturer, Bristol University, 1938-39, instructor, University of Texas, 1939-41; University of Iowa, 1941-46. Socony-Vacuum Laboratories, 1946-49. Bell Telephone Laboratories, 1949-. Mr. Wannier, a theoretical physicist in the Physical Research Department, has worked on photoconductivity and related phenomena, and the motion of ions in gases. Also Mission to Germany, 1945. Member of American Physical Society and Schweizer Physikalische Gesellschaft.

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXII

MARCH 1953

NUMBER 1

KANSAS CITY
PU: 1
MAR 30 1953

Ferrite Core Inductors	H. A. STONE, JR.	265
A Throwdown Machine for Telephone Traffic Studies	G. R. FROST, WILLIAM KEISTER AND ALISTAIR E. RITCHIE	292
Working Curves for Delayed Exponential Calls Served in Random Order	ROGER I. WILKINSON	360
Magnetic Resonance: Part II—Magnetic Resonance of Electrons	KARL K. DARROW	384
A Study of Non-Blocking Switching Networks	CHARLES CLOS	406
The Evaluation of Wood Preservatives—Part II	REGINALD H. COLLEY	425
Abstracts of Bell System Papers Not Published in this Journal		506
Contributors to this Issue		519

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

S. BRACKEN, *President, Western Electric Company*

F. R. KAPPEL, *Vice President, American Telephone
and Telegraph Company*

M. J. KELLY, *President, Bell Telephone Laboratories*

EDITORIAL COMMITTEE

E. I. GREEN, *Chairman*

A. J. BUSCH

F. R. LACK

W. H. DOHERTY

J. W. MCRAE

G. D. EDWARDS

W. H. NUNN

J. B. FISK

H. I. ROMNES

R. K. HONAMAN

H. V. SCHMIDT

EDITORIAL STAFF

PHILIP C. JONES, *Editor*

M. E. STRIEBY, *Managing Editor*

R. L. SHEPHERD, *Production Editor*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; Alexander L. Stott, Treasurer. Subscriptions are accepted at \$3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXII

MARCH 1953

NUMBER 2

Copyright, 1952, American Telephone and Telegraph Company

Ferrite Core Inductors

By H. A. STONE, JR.

(Manuscript received November 19, 1952)

This paper describes the use of ferrite materials as cores for inductors and develops methods for taking maximum advantage of their properties in the design of inductors for communication circuits.

INTRODUCTION

The extent to which the theoretical capabilities of wave filters and networks can be realized in practice usually depends on how high a Q , ratio of reactance to resistance, can be obtained in the inductors. In the voice and carrier telephone frequency ranges the dissipation in mica capacitors is so small compared to that in inductors that it can generally be neglected. Even paper capacitors, in the lower frequency ranges, compare favorably in Q with the best available coils. Consequently, there has been considerable incentive to develop improved magnetic materials that will permit the realization of higher Q inductors for filter and network use.

Work along this line has resulted in the development of the permalloys, and later the molybdenum permalloys which, powdered, insulated and pressed into shapes, have become the standard materials for wave filter coils in the voice and carrier frequency ranges.^{1, 2}

Although the permalloys represent a vast improvement over the soft iron that preceded them they share the fundamental disadvantage of all metals, that they are good conductors. Any conductor in the vicinity of an alternating magnetic field has eddy currents induced in it, and, if the conductor is the core of a coil, the power loss associated with these currents appears as added resistance in the coil windings. To restrict the eddy current paths it is customary to powder the material and insulate the particles with some kind of inorganic filler. However, there

are two practical limitations on this procedure. First, there is the mechanical difficulty of grinding particles to smaller than the few microns diameter that now represents the commercial limit. Second, as the size of the individual particles is made smaller more air gaps are introduced in the magnetic circuit by the insulating material, and the effective over-all permeability of the core is reduced. If carried too far the advantages in using the magnetic material in the first place are largely lost.

The need for high resistivity magnetic materials has been recognized for many years, and, in fact, the naturally occurring magnetic iron oxide, Fe_3O_4 , or lodestone, is such a material. Its permeability, however, is only about 3 or 4 and its use in coil work has been very restricted. The problem of producing an oxide or mixture of oxides having higher permeabilities was investigated by Philips Gloeilampenfabrieken in Holland during the late war and they were successful in producing oxide mixtures, or ferrites, having permeabilities of 1500 and higher.³

After the war, Bell Telephone Laboratories initiated a program of ferrite development, and at the present time ferrite cores as well as inductors and transformers which take advantage of their unique properties are in commercial production.

The most commonly used ferrites consist of solid solutions of oxides of manganese, zinc and iron, or nickel, zinc and iron. They are prepared by mixing the materials, pressing them into the required shape, and heat treating them under carefully controlled conditions. The resulting ferrite parts are hard and brittle but they can be machined with diamond tools.

Although ferrites belong in the class of so-called semiconductors their conductivities are of the order of a millionth of those of metals and the eddy current losses are proportionately smaller. There are frequency limitations in ferrites but these occur at higher frequencies than are ordinarily of concern in magnetic core inductor work.

Besides offering the coil designer new possibilities because of low eddy current losses the ferrites have a secondary advantage in that they do not have to be subdivided to keep those losses low. The powdered metallic materials are fragile, are difficult to press except in simple shapes such as rings or cylindrical plugs, and are difficult to machine. As a result the coil designer, for virtually every use, is restricted to the toroidal type coil which, although it has some important advantages, is inherently expensive to wind and adjust. With ferrites, on the other hand, a variety of shapes can be produced. They have ample strength and can be ground and machined.

The combined magnetic and mechanical advantages of ferrites put

them in an entirely new class as core materials for inductors. To exploit these advantages properly it is necessary to reconsider some of the fundamental aspects of coil design, since some of the assumptions that applied to powdered core coils are no longer valid, and others have to be modified.

DISSIPATION FACTOR

Legg has shown that the core loss characteristics of metallic magnetic materials can be described by three constants, defined as eddy current (e), hysteresis (a), and residual (c), coefficients, respectively.¹ The total increment of resistance (R_m), due to core losses is given by

$$R_m = e\mu f^2 L + a\mu B_m f L + c\mu f L \quad (1)$$

where μ = effective permeability of the core structure

f = frequency in c.p.s.

L = inductance in henrys

B_m = max. flux density in gauss

This expression is valid for practical applications so long as the flux density is low.

Measurements on ferrite cores indicate that the same formula can be used, with appropriate constants, at least for frequencies up to 200 or 300 kc, and, again, for low flux density applications. However, the emphasis on the terms becomes quite different. The residual loss, as its name implies, is usually of negligible importance in metallic magnetic materials, but in ferrites because of the very low value of the eddy current constant, the residual loss often emerges as a controlling factor.

Beside the core losses the other factors in an inductor that contribute to the total measured resistance are the dc resistance of the winding, ac losses in the winding and the parasitic capacitances within the structure. The latter two, especially the capacitance effects, have always had to be taken into account in higher frequency work, but their contribution to the total loss becomes a matter of first order importance when ferrites are used since they are no longer small in comparison with the eddy current loss in the core.

The objective in the design of inductors for wave filter and similar applications is to provide a specified inductance with as low an associated resistance as practicable. The quality factor, Q , the ratio of reactance to resistance, is the traditional measure of the extent to which this has

been achieved. In the discussion that follows, however, it will be more convenient to use the reciprocal of Q as the measure of quality. This permits combining the loss components in a simple additive manner. Thus

$$\text{Dissipation Factor} = D = \frac{1}{Q} = \frac{R_{dc} + R_e + R_h + R_r + R_c + R_s}{\omega L} \quad (2)$$

where R_{dc} = dc resistance of the winding

R_e = eddy current loss

R_h = hysteresis loss

R_r = residual loss

R_c = increment of measured resistance due to distributed capacitance

R_s = ac loss in wire of the winding.

or

$$D = D_{dc} + D_e + D_h + D_r + D_c + D_s \quad (3)$$

where $D_n = \frac{R_n}{\omega L}$ (n being any of the above subscripts).

Having been given the inductance, and the operating frequency and current, for a desired inductor, and having chosen the core material that will be used, each of the D 's above will be functions of the following variables: Permeability, Volume and Proportions of the core structure. We will now investigate how each of these factors can be manipulated to yield the highest Q (or lowest D) coil.

EFFECT OF PERMEABILITY ON DISSIPATION FACTOR

The permeability of a permalloy powder core is determined by the fineness of the powder and the amount of insulation with which it is mixed before firing. Obviously it is commercially impracticable to manufacture cores having a great variety of permeabilities, and it has happened that four values have been standardized for commercial use; 125, 60, 26 and 14. The coil designer working with permalloy powder, beyond choosing the most appropriate of these four values, seldom finds it practicable to control the permeability. This is due to the difficulty of pressing shapes suitable for use with air gaps, and the fragility of the parts. With ferrite, on the other hand, the parts can readily be adapted to the control of permeability by insertion of air gaps, and thus establish-

ment of the optimum permeability becomes an important step toward achieving the desired coil performance.

The first four of the "D's" in (3), dc resistance and the core losses, are direct functions of the permeability and can be discussed to a certain extent independently of the last two.

The inductance of a magnetic core inductor is given by

$$L = \frac{kN^2 A \mu}{\ell} \quad (4)$$

where $k = \text{constant}$

$N = \text{number of turns in winding}$

$A = \text{cross section of core}$

$\ell = \text{mean length of magnetic path through core.}$

If the geometric proportions of the core have been prescribed, A and ℓ will bear a constant relationship to $V^{2/3}$ and $V^{1/3}$, respectively, and (4) may be written

$$L = k_1 N^2 V^{1/3} \mu \quad (5)$$

where $k_1 = \text{constant}$

$V = \text{core volume.}$

The dc resistance in the winding is

$$R_{dc} = \frac{\rho N^2 \lambda}{k_w W} \quad (6)$$

where $\rho = \text{resistivity of the conductor}$

$\lambda = \text{mean length of turn}$

$W = \text{available area of cross section through which turns can be linked with the core}$

$k_w = \text{winding efficiency. This is the ratio of the actual cross section of conductor to the available area, } W.$

Again, for a core of given proportions (6) may be written

$$R_{dc} = \frac{k_2 N^2}{V^{1/3}} \quad (7)$$

Eliminating N from (5) and (7)

$$R_{dc} = \frac{k_3 L}{V^{2/3} \mu} \quad (8)$$

and

$$D_{dc} = \frac{k_4}{V^{2/3} \mu f} \quad (9)$$

where k_2, k_3, k_4 are constants.

From (1) and (2) it is seen that

$$D_e = k_5 \mu f \quad (10)$$

$$D_h = k_6 B_m \mu = \frac{k_{10} \mu^{3/2} L^{1/2} i}{V^{1/2}} \quad (11)$$

$$D_r = k_7 \mu \quad (12)$$

in which the flux density,

$$B_m = \sqrt{2} \mu H = k_8 \mu \frac{Ni}{\ell} = k_9 \sqrt{\frac{\mu L i^2}{\ell A}} = k_9 \sqrt{\frac{\mu L i^2}{V}}. \quad (13)$$

H is the field strength due to the r.m.s. current, i , and k_5 to k_{10} are constants.

Combining (9) through (12)

$$D = \frac{k_4}{V^{2/3} \mu f} + k_5 \mu f + k_{10} \frac{\mu^{3/2} L^{1/2} i}{V^{1/2}} + k_7 \mu. \quad (14)$$

For a specified inductance at a given frequency and current the optimum permeability can be found from (14) by inserting the numerical values and solving graphically. It is very unlikely, however, that this cumbersome procedure would be necessary in practical work. Usually, depending on the design requirements, one or another of the core loss factors will be comparatively large and the others can be neglected. We will examine separately the three cases where residual, hysteresis and eddy current loss, respectively, predominate over the other core losses, and show how the permeability can be adjusted to minimize the dissipation factor.

1. DC Resistance and Residual Loss Predominate

This condition, which formerly was very seldom encountered, is becoming of increasing practical importance for two reasons. First, the eddy

current losses in ferrite are extremely small. Second, the introduction of the transistor has created many new applications for inductors for use at very low power levels, and with proportionately lower hysteresis losses. For this condition we have, from (14)

$$D = \frac{k_4}{V^{2/3}\mu f} + K_7\mu \quad (15)$$

Solving for the value of μ , $\mu_{\text{opt.}}$, which will yield the lowest dissipation

$$\frac{dD}{d\mu} = -\frac{k_4}{V^{2/3}\mu^2 f} + k_7 \quad (16)$$

$$\mu_{\text{opt.}} = \frac{k_4^{1/2}}{k_7^{1/2} V^{1/3} f^{1/2}}$$

and

$$D_{\text{opt.}} = 2 \frac{k_4^{1/2} k_7^{1/2}}{V^{1/3} f^{1/2}} \quad (17)$$

It may be noted from (16) and (17) that the optimum permeability is that which will result in the dc resistance and residual loss being equal.

2. DC Resistance and Hysteresis Loss Predominate

This condition, while it does not often apply to conventional size inductors for transmission networks, may occur when miniaturization of coils is contemplated. It will be noted from (14) that hysteresis is the only core loss component that is dependent on the volume of the core. We have, from (14)

$$D = \frac{k_4}{V^{2/3}\mu f} + \frac{k_{10}\mu^{3/2}L^{1/2}i}{V^{1/2}} \quad (18)$$

The optimum permeability and dissipation factor derived from this equation are

$$\mu_{\text{opt.}} = \sqrt[5]{\frac{4k_4^2}{9k_{10}^2 f^2 L i^2 V^{1/3}}} \quad (19)$$

$$D_{\text{opt.}} = [(3/2)^{2/5} + (2/3)^{3/5}] \sqrt[5]{\frac{k_4^3 k_{10}^2 L i^2}{f^3 V^3}} \quad (20)$$

Inspection shows that

(a) The optimum permeability is such that the ratio of dc resistance to hysteresis loss is 3/2.

(b) This value will depend not only on the frequency and volume of core, but on the inductance and current level as well.

3. DC Resistance and Eddy Current Loss Predominate

This condition is the most commonly occurring of all when metallic magnetic materials are used but is of very little importance for ferrite coil design. From (14)

$$D = \frac{k_4}{V^{2/3}\mu f} + k_5\mu f \quad (21)$$

$$\mu_{\text{opt.}} = \frac{k_4^{1/2}}{k_5^{1/2}V^{1/3}f} \quad (22)$$

$$D_{\text{opt.}} = \frac{2k_4^{1/2}k_5^{1/2}}{V^{1/3}}. \quad (23)$$

Here we note that, as in the case for residual loss, optimum permeability is that which assures that the core loss is equal to the dc resistance. If both eddy current and residual losses were significant the sum of these should be equal to the DC resistance. It is also seen from (22) and (23) that although the optimum permeability is a function of frequency the resulting dissipation factor is independent of the frequency.

We have developed expressions for optimum permeability for three conditions of core loss. It is now of interest to know how critical the adjustment of permeability is. Fig. 1 shows the effect on the dissipation factor of deviations from optimum in permeability. It is seen that for

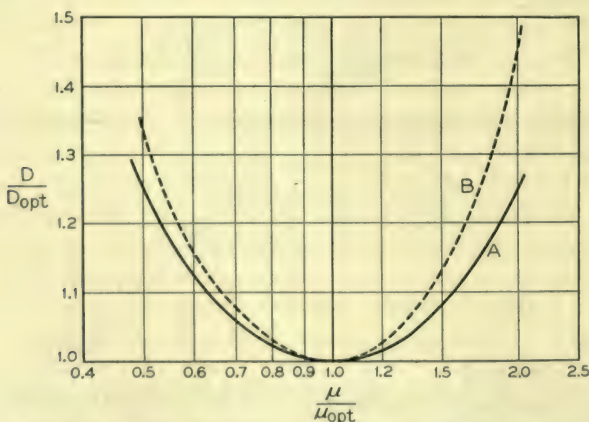


Fig. 1 — Effect of permeability on dissipation factor.

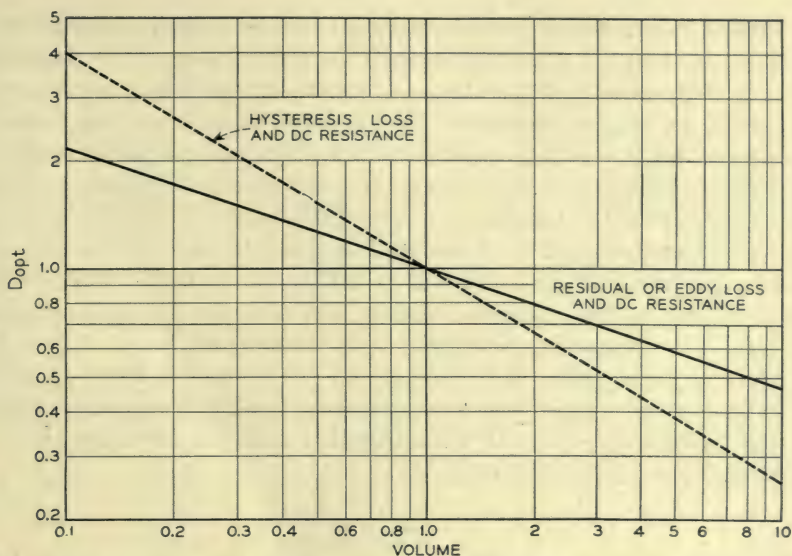


Fig. 2 — Relationship between core volume and dissipation factor.

deviations from optimum up to about 20 per cent the penalty in dissipation is not severe, but beyond this point the dissipation factor increases rapidly.

EFFECT OF VOLUME ON DISSIPATION FACTOR

It has been shown in the preceding section that the permeability can be adjusted for optimum coil quality but that the value to which it is adjusted depends on which of the loss factors predominates. It will also be seen from the foregoing that both the value of optimum permeability and the dissipation factor that can be realized by so adjusting the permeability depend on the volume of the core. The curves of Fig. 2 are derived from (17), (20) and (23), respectively, considering D_{opt} as a function of the independent variable, V . That is, they show how the dissipation factor will vary as a function of volume assuming that the permeability is adjusted as the volume is changed so that the lowest possible dissipation factor is always realized.

As would be expected, these curves show that, regardless of the relative magnitudes of the core losses, the dissipation factor can be decreased by increasing the size of the coil. There are serious limitations, however, on how far this process can be carried, and these are due to D_c and D_s , the resistive components resulting from capacitance and ac loss in the wire.

This ac loss is due to a combination of skin effect and eddy currents in the conductor. It is proportional to the sixth power of the diameter of each of the strands of wire that go to make up the conductor, and the square of the number of strands⁴. If the size of an inductor is increased, therefore, the ac loss may rapidly assume important proportions. Some help can be obtained by finer stranding of the wire and this has, in some cases, been carried to the extent of using 810 separately insulated strands in a single conductor. Even if fine stranding is sufficient to reduce the ac loss to a tolerable value, a penalty is paid in the form of higher dc resistance, due to the space occupied by the separate insulation on the strands. This amounts to a decrease in the winding efficiency, k_w , in (6).

The effect of distributed capacitance on the dissipation factor of a coil is a function not only of the volume, directly, but of the absolute value of the dissipation factor. If the dissipation factor is low compared to unity it is given, approximately, by

$$D_c = \frac{(D_0 + d) \frac{C}{C_g}}{1 - \frac{C}{C_g}} \quad (24)$$

where $D_0 = D - D_c$ is the dissipation factor that would obtain were it not for capacitance

d = dissipation factor of the distributed capacitance

C = distributed capacitance

C_g = resonating capacitance of the inductor.

If (24) is written in an alternative form

$$\frac{D}{D_0} = \frac{1 + \frac{d}{D_0} \frac{C}{C_g}}{1 - \frac{C}{C_g}} \quad (25)$$

it becomes evident that a coil with an inherently high Q is more seriously affected, proportionately, by distributed capacitance than is a low Q coil. Fig. 3 shows this information graphically, using the value 0.01 for d , which measurements on model inductors indicate to be of an appropriate magnitude.

Each of the curves in Fig. 3 is based on a constant value for C/C_g . In practice, however, as volume is increased it becomes more difficult to

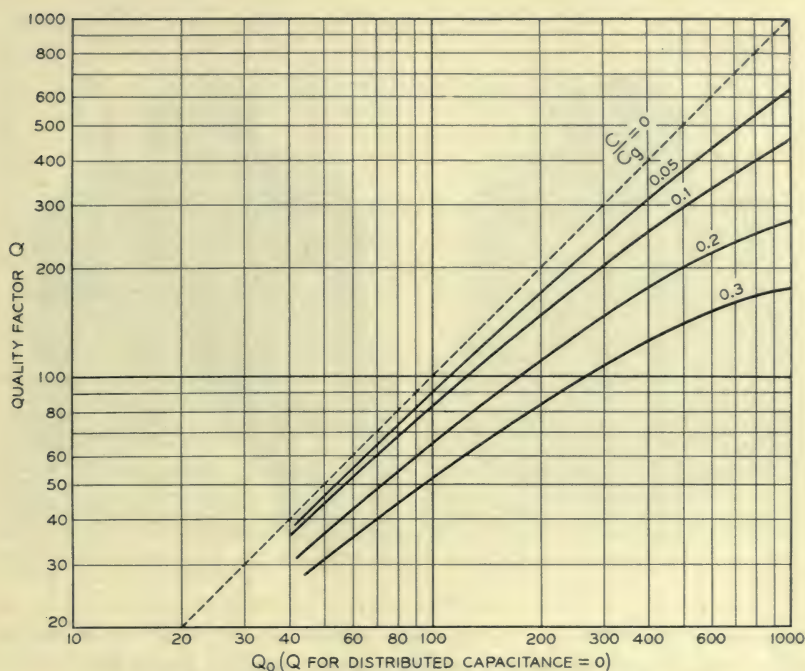


Fig. 3 — Effect of distributed capacitance on Q .

maintain small values of distributed capacitance. It will be necessary to increase the separation of windings from each other and from the core and this will result in a lower winding efficiency and a higher dc resistance.

Since the limitations imposed by ac losses in the wire and by distributed capacitance depend on the absolute magnitudes of the design parameters and the mechanical details of the inductor structure they do not lend themselves to representation in practicable generalized formulas as do the core losses. However, the following information on some model inductors will illustrate the magnitudes of these limitations.

A ferrite core coil was constructed similar to that shown in Fig. 4, but having a core volume of only 0.04 cubic inches. It was a 5 mh coil for use at 100 kc and had a distributed capacitance of 10 mmf. It was wound with a single strand conductor and the winding efficiency, k_w , was 0.4. Since it was intended for use at very low power levels the hysteresis loss was negligible. The permeability was optimized in accordance with case 1, above, for residual loss predominating. The measured Q was very close to 300. Thus, $D = 0.0033$. From the above data we note that $C/C_g = 0.02$, and from (25) we can calculate that $D_0 = 0.0030$,

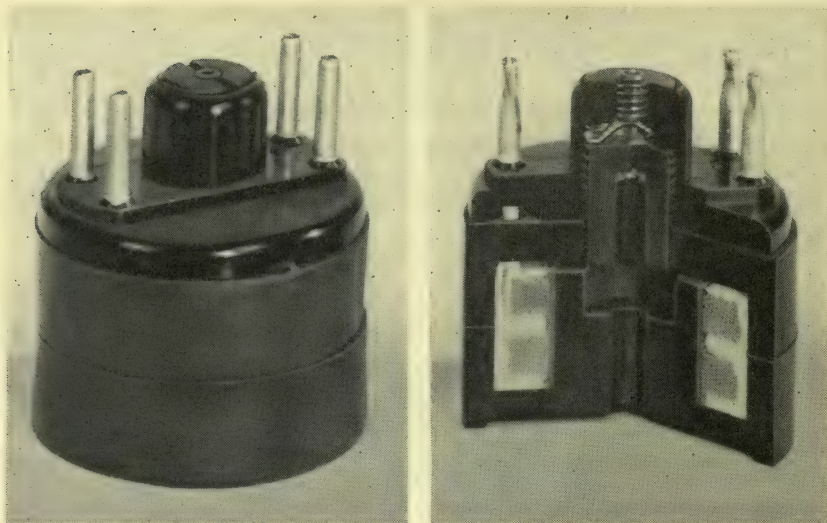


Fig. 4 — The 1509 type adjustable ferrite inductor for Type-O carrier system filters.

using the value 0.01 for d . Thus the “geometric Q ”, that which would obtain were it not for the distributed capacitance, is about 330.

The 1509 type inductor, shown in Fig. 4, was designed for use in the type-O carrier system.⁵ It has a volume of 1.4 cubic inches, 35 times that of the small coil discussed above. From the standpoint of dc resistance and core loss alone, (7) indicates that the dissipation factor should be

$$D_0 = \frac{1}{35^{1/3}} (0.0030) = 0.00092,$$

or that the Q should be greater than 1000.

In order to keep the capacitance to the same order as that of the small coil, and to use stranded wire to avoid excessive ac loss, the winding efficiency, k_w , had to be reduced from 0.4 to 0.13. The effect of this on dissipation can be shown from (17), remembering that k_4 varies inversely as k_w :

$$\frac{D'_0}{D_0} = \frac{(3k_4)^{1/2}}{k_4^{1/2}} = 1.73$$

$$D'_0 = (1.73)(0.00092) = 0.00159.$$

Thus, the Q (still neglecting the effect of the 10 mmf distributed capacitance that remains after reducing the winding efficiency) should be 630.

From (24)

$$D_c = \frac{(0.00159 + 0.01)0.02}{0.98} = 0.00024$$

and

$$D = D_0 + D_c = 0.00183$$

or the actual measured Q of the large inductor is 550, about half that indicated by core loss considerations alone.

EFFECT OF CORE PROPORTIONS ON DISSIPATION FACTOR

In the foregoing discussion of coil volume it has been assumed that the core proportions remained fixed as the volume was changed. It is now of interest to know what these proportions should be to insure the lowest dissipation.

The general type of structure under consideration consists of a closed cylindrical container and a center post of magnetic material, and an air gap which might be anywhere in the magnetic path. It is assumed that the thickness of the shells is such that the area of the magnetic path is uniform and equal to the area of cross section of the post. It is also assumed that the flux is uniformly distributed within the cross-section of the core. Fig. 5 represents such a core schematically.

Given a fixed over-all coil volume it is desired to know what proportions should apply to the outside diameter, the diameter of the center post and the axial height of the structure, to provide the lowest dissipation factor. This will be examined first for the case where residual or eddy

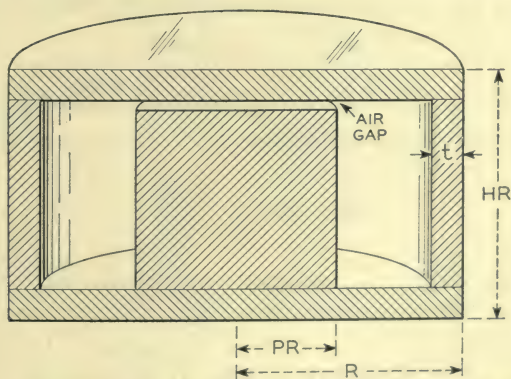


Fig. 5 — Optimum proportions for post and shell core assembly.

current losses are large compared to hysteresis loss, and then for the situation where hysteresis loss predominates. Referring to Fig. 5, the following relationships are noted.

$$t \text{ (thickness of shell) is given by } (R - t) = R\sqrt{1 - P^2} \quad (26)$$

$$A \text{ (magnetic cross-section) } = \pi P^2 R^2 \quad (27)$$

$$\ell \text{ (mean magnetic path) } = R(3\sqrt{1 - P^2} - 1 + 2H) \quad (28)$$

$$W \text{ (available winding cross-section)} \\ = R^2(\sqrt{1 - P^2} - P)(2\sqrt{1 - P^2} - 2 + H) \quad (29)$$

$$\lambda \text{ (mean length of turn) } = \pi R(\sqrt{1 - P^2} + P) \quad (30)$$

$$V_c \text{ (coil volume) } = \pi R^3 H \quad (31)$$

1. DC Resistance, and Residual or Eddy Current Losses, Predominate

The dissipation factor, due to the dc resistance, is seen from (4) and (6) to be

$$D_{dc} = \frac{R_{dc}}{2\pi f L} = k_{11} \frac{\lambda \ell}{WA\mu} \quad (32)$$

where $k_{11} = (\rho/2\pi k k_w)$ is a constant with respect to the core proportions, assuming the winding efficiency is not materially affected by changes in shape.

Since, regardless of shape, we will want to adjust the air gap so that μ is optimum, and since we know from the previous discussion that this will occur when the dc resistance is equal to the sum of the core losses, we have, from (10) and (12)

$$D_{dc} = D_e + D_r = (k_5 f + k_7) \mu \quad (33)$$

Eliminating μ from (32) and (33):

$$D_{dc} = k_{12} \sqrt{\frac{\lambda \ell}{WA}} \quad (34)$$

where $k_{12} = \sqrt{k_{11}(k_5 f + k_7)}$ is again a constant with respect to the core dimensions. The dissipation factor is

$$D = 2D_{dc} = 2k_{12} \sqrt{\frac{\lambda \ell}{WA}} \quad (35)$$

Putting in the values, from (27) to (30), for λ , ℓ , W and A :

$$D = \frac{2k_{12}}{R} \sqrt{\frac{(\sqrt{1-P^2}+P)(3\sqrt{1-P^2}-1+2H)}{P^2(\sqrt{1-P^2}-P)(2\sqrt{1-P^2}-2+H)}} \quad (36)$$

Since the volume of the coil, given in (31), is assumed to be constant, we can eliminate R from the above equation:

$$D = k_{13} \sqrt{\frac{(\sqrt{1-P^2}+P)(3\sqrt{1-P^2}-1+2H)H^{2/3}}{P^2(\sqrt{1-P^2}-P)(2\sqrt{1-P^2}-2+H)}} \quad (37)$$

$$\text{where } k_{13} = \frac{2\pi^{1/3}k_{12}}{V_c^{1/3}}$$

We now have an expression for the dissipation factor in terms of the two variables, P and H , which determine the proportions of the coil structure. The effects of these proportions on the dissipation factor are shown in Fig. 6. It will be seen that best results are achieved when the radius of the post is approximately 0.45 of the outside radius and the axial height is about 1.2 times this radius. Fig. 5 is drawn to this scale, and the 1509 type coil shown in Fig. 4 approximates these proportions.

2. DC Resistance and Hysteresis Loss Predominate

We have noted that for this case optimum permeability is that which results in the following relationship between the dc and hysteresis dissipations:

$$3D_h = 2D_{dc} \quad (38)$$

From (11) and (13)

$$D_h = k_6 B_m \mu = \frac{k_6 k_9 \mu^{3/2} L^{1/2} i}{\ell^{1/2} A^{1/2}} \quad (39)$$

Putting the values from (32) and (39) in (38), we can eliminate μ and express D_{dc} in terms of the dimensional variables of the coil:

$$D_{dc} = k_{14} \frac{\lambda^{3/5} \ell^{2/5}}{W^{3/5} A^{4/5}} \quad (40)$$

$$\text{where } k_{14} = \left(\frac{3}{2} k_6 k_9 k_{11}^{3/2} L^{1/2} i\right)^{2/5}$$

The total dissipation factor is

$$D = D_{dc} + D_h = \frac{5}{3} D_{dc} = \frac{5}{3} k_{14} \frac{\lambda^{3/5} \ell^{2/5}}{W^{3/5} A^{4/5}} \quad (41)$$

Using (27) to (30) we can express D in terms of the coil proportions

$$D = \frac{5}{3}k_{14} \sqrt[5]{\frac{(\sqrt{1-P^2} + P)^3(3\sqrt{1-P^2} - 1 + 2H)^2}{R^3P^8\pi(\sqrt{1-P^2} - P)^3(2\sqrt{1-P^2} - 2 + H)^3}} \quad (42)$$

Again, eliminating R by use of the constant volume relationship, (31)

$$D = k_{15} \sqrt[5]{\frac{H^3(\sqrt{1-P^2} + P)^3(3\sqrt{1-P^2} - 1 + 2H)^2}{P^8(\sqrt{1-P^2})^3(2\sqrt{1-P^2} - 2 + H)^3}} \quad (43)$$

where $k_{15} = \frac{5\pi^{2/5}k_{14}}{3V_c^{3/5}}$

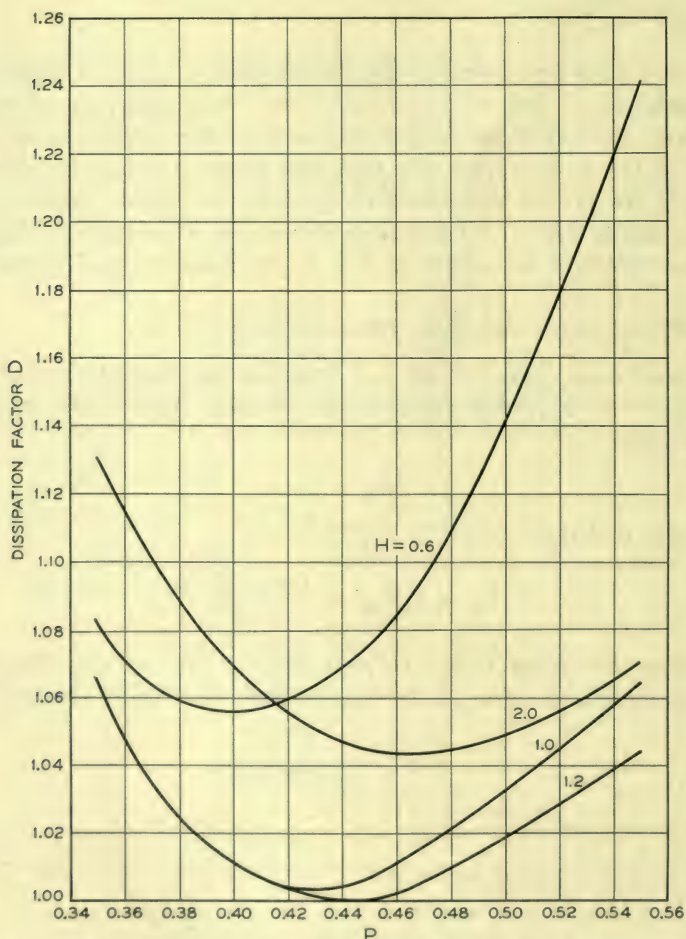


Fig. 6 — Effect of core proportions on dissipation factor when hysteresis loss can be neglected.

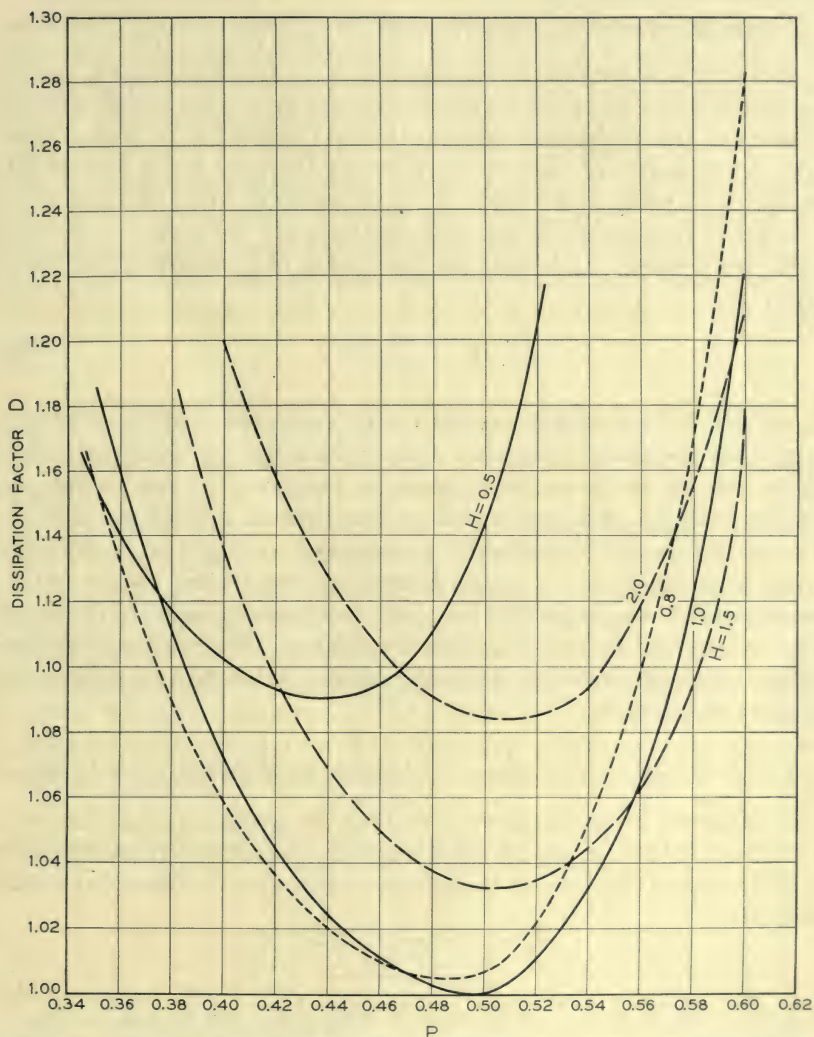


Fig. 7 — Effect of core proportions on dissipation factor when hysteresis loss is predominate.

This information is shown graphically in Fig. 7. It will be seen that when the hysteresis losses are important both the post diameter and the axial height should be about half of the overall diameter. These proportions are not far different from those derived as optimum from (37). This is fortunate since it means that cores of the same proportions will be suitable in different applications regardless of which core losses predominate.

METHODS OF INDUCTANCE ADJUSTMENT

The fact that ferrites can be molded in a variety of shapes, can be machined, and can be used with discrete air gaps permits considerable freedom in the mechanical methods that can be used to provide continuous adjustability of inductance. Whether the need is for factory adjustment to prescribed values or for adjustment after the coil is assembled in its equipment, the basic methods are the same.

We have noted, in (4) that the inductance of a magnetic core coil is given by

$$L = \frac{kN^2 A \mu}{\ell} \quad (4)$$

in which k is a constant depending on the units used, and the other four quantities are design variables. Any of these, or any combinations of them, can be manipulated to produce variations in the inductance. They are: length of magnetic path, ℓ ; cross section of magnetic path, A ; number of turns, N ; and effective permeability, μ . There is an additional variable hidden by the formula's assumption that perfect coupling exists among all the turns in the winding. Inductance can be adjusted by changing the coupling between parts of the winding. Each of these five variables, which are of widely differing practical value, will be commented upon at least briefly.

1. Adjustment of Inductance by Change in Magnetic Path Length, ℓ

Adjustment by a change in ℓ without an accompanying change in μ requires that the air gap also be changed since the effective permeability is a function of the ratio of the air gap dimensions to those of the total structure.

$$\mu = \frac{\mu_m}{1 + \mu_m \frac{g}{a}} \quad (44)$$

where μ_m = permeability of the core material

g = ratio of the length of air gap to total length, ℓ

a = ratio of effective cross section of gap to core cross section, A .

It would be mechanically possible to design an adjustable inductor of this sort but it is unlikely that there would be any practical advantage in maintaining constant permeability over the adjusting range. On the

contrary, it would probably be more advantageous as well as mechanically simpler to have the air gap length constant so that the increase in inductance due to shortening the magnetic path would be augmented by an increase in permeability. In a device such as shown schematically in Fig. 8, as the two windings approach each other the over-all magnetic path decreases and at the same time the area of the air gap increases. Actually, the preponderant effect is due to the air gap change and the effect of variation in length of path becomes of secondary importance. This is likely to be the case generally, when both path length and air gap change simultaneously, since most of the reluctance is in the air gap.

2. Adjustment of Inductance by Change in Magnetic Cross Section, A

To provide for change of inductance by constriction of a part of the magnetic cross section is apt to be undesirable since it forces a concentration of flux in the constricted part of the magnetic circuit and may introduce unduly high hysteresis losses. Even if the levels are low enough so that this is not a consideration the amount of inductance variation that can be achieved even by a large constriction in part of the core is relatively small. To illustrate this we will consider a structure such as shown schematically in Fig. 9, in which one sector of the core can be varied in effective cross section. The reluctance of the structure is equal to the sum of the reluctances of the fixed part of the core, the sector of length $n\ell$, whose cross section aA can be varied, and the air gap:

$$\begin{aligned} R &= \frac{(1-n)\ell}{\mu_m A} + \frac{n\ell}{\mu_m aA} + \frac{g\ell}{A} \\ &= \frac{\ell}{\mu_m A} \left[1 + n \left(\frac{1}{a} - 1 \right) + g\mu_m \right]. \end{aligned} \quad (45)$$

Let us assume the arbitrary but reasonable values of 2000 for μ_m , 0.01 for g , and 0.1 for n . Then approximately

$$R = \frac{\ell}{2000A} \left(20 + \frac{0.1}{a} \right).$$

It will be seen that as a is varied from 1.0, corresponding to the full cross section of the main core, to one-tenth of that, the change in inductance, which is inversely proportional to reluctance, will only amount to about 5 per cent.

3. *Adjustment of Inductance by Change in Number of Turns, N*

An adaptation of turns adjustment for use in a shell type structure is shown in Fig. 10.⁶ The center post with the winding on it can revolve and turns can be removed by pulling on the outer lead, or added by rotating the knob at the end of the shaft. Continuous adjustment is possible since it is not necessary to add or remove integral numbers of turns. Although an inductor of this sort involves some mechanical complexity it has the advantage that the core parts do not have to be precisely machined. The turns adjustment can be used to compensate for sizable variations in the dimensions of the core parts and the resulting air gaps.

4. *Adjustment of Inductance by Change in Permeability, μ*

Adjustment by change in permeability, that is, change in length or area of air gaps, can be accomplished in a variety of ways to meet differing design needs and can be mechanically simple and economical. For most purposes permeability adjustment will offer more advantages than any of the other methods.

Whatever the method of adjustment used its effectiveness will depend on proper correlation between the mechanical motion that produces the change and the inductance itself. For most filter and network applications the following considerations will apply:

(a) The slope of the line showing inductance plotted against the displacement that produces the adjustment should be reasonably constant over the adjusting range.

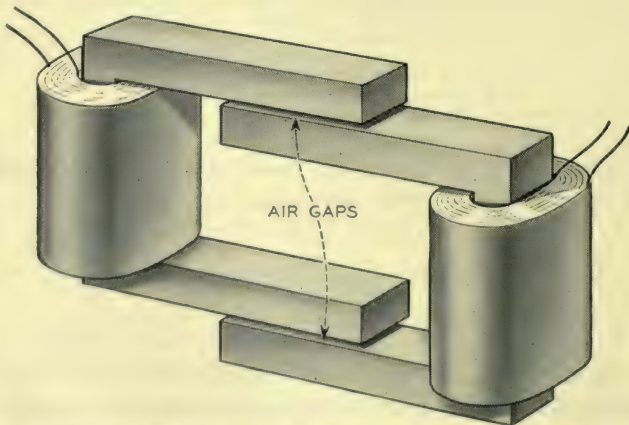


Fig. 8 — Inductance adjustment by decrease in magnetic path length and increase in air gap area.

(b) The slope should not be negative at any position in the adjusting range as this would introduce the possibility of false or double balance when the coil is being adjusted by null or peaking methods.

(c) The slope should not be so small that play in the mechanical parts will cause changes in slope of the same magnitude as the slope itself.

(d) Within the above limitation the smaller the slope the more precisely the adjustment can be made.

(e) Conversely, the greater the slope the greater the range of adjustment for a given amount of mechanical motion.

(f) It follows from (d) and (e) that the amount of mechanical motion available determines the product of range and precision. Where the mechanical motion is rotary, such as with adjustment by turning a screw, it is possible to adjust a coil to a precision of about $1/400$ of a revolution without undue difficulty. If the range is covered by N revolutions of the adjusting screw, and the over-all range is $\pm R$ per cent of the mean value:

$$P = \frac{2R}{400N} = \frac{R}{200N}$$

where P = the precision of adjustment in per cent of the nominal value.

In the coil shown in Fig. 4, six turns of the adjusting screw are effective in producing an over-all change of ± 15 per cent in the inductance. The precision with which the adjustment can be made is, therefore, very close to ± 0.01 per cent of the value desired.

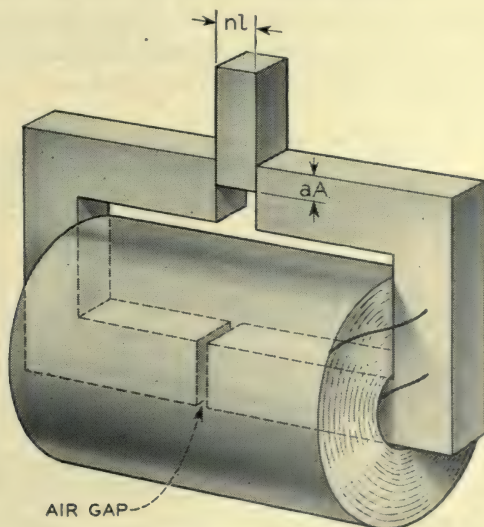


Fig. 9 — Inductance adjustment by constriction in magnetic cross section.

(g) It would appear from the above that by gearing or other means of increasing the mechanical motion that governs the inductance change it should be possible to improve the precision. This is true up to the point where consideration (c) is violated, to the point where play in the mechanical parts permits variations of the same order as the nominal precision.

An obvious form of permeability adjustment would consist of a means for varying the distance between two plane magnetic surfaces, as shown schematically in Fig. 11 (a). One disadvantage of this, or of any other means that involves a change in air gap length, is in the nonlinearity of adjustment. The effective permeability of a coil is given in (44). In most voice and carrier frequency applications of ferrite the magnitudes of g and a will be such that, approximately

$$\mu = \frac{a}{g} \quad (46)$$

and, if the cross section of the air gap is the same as that of the core

$$\mu = \frac{1}{g}.$$

A typical adjustment curve resulting from this inverse relationship is shown in Fig. 11 (b).

In addition to its nonlinearity the simple butted gap has a shortcoming in that the mechanical motions involved are of the order of only a few hundredths of an inch, which requires that parts be very accurately fitted. This can be somewhat alleviated by using cone or wedge shaped

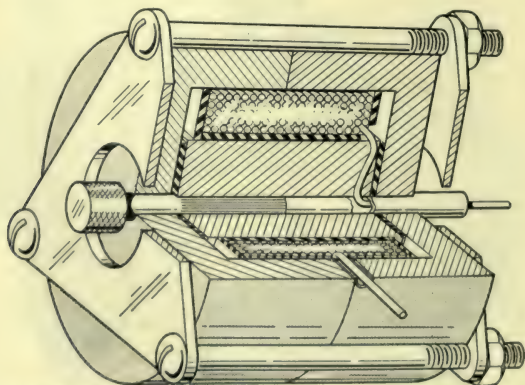


Fig. 10 — Inductance adjustment by addition or removal of turns.

gaps, as illustrated in Fig. 11 (c). Here the distance d , travelled by the screw is longer than the effective air gap g , by the amount

$$d = \frac{g}{\sin \theta}.$$

One means of overcoming the nonlinear characteristic of gaps such as these is to introduce compensation in the form of a secondary gap that opens as the main gap closes.⁷ Such an arrangement is shown in Fig. 12 (a). When there are two gaps in series their reluctances are additive and their effect on permeability is given approximately by

$$\mu = \frac{1}{g_1 + g_2}. \quad (47)$$

Fig. 12 (b) shows the inductance characteristics that would result from either of the two gaps alone, and the effect of the two gaps in series.

From (46) it can be seen that the effective permeability varies ap-

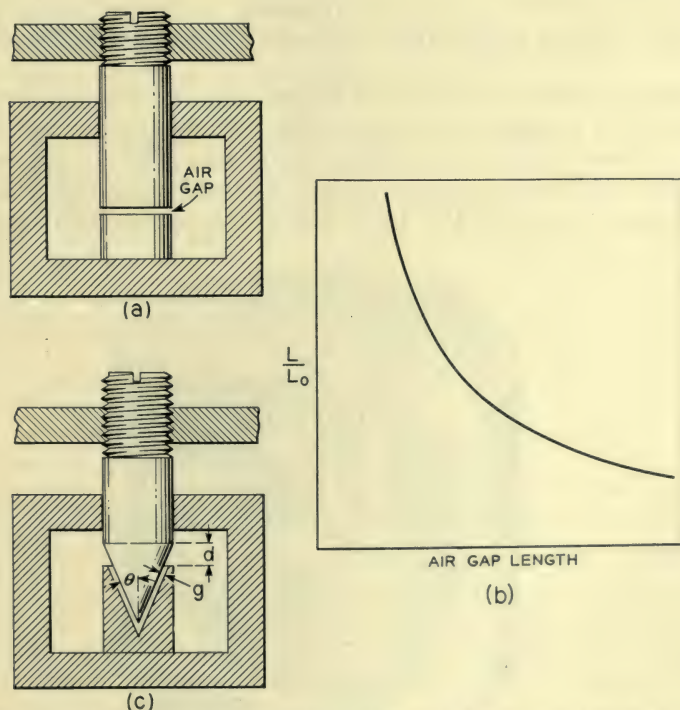


Fig. 11 — Inductance adjustment by variation of air gap length. (a) Gap formed by parallel plane surfaces. (b) Adjustment characteristic. (c) Cone or wedge shaped gaps.

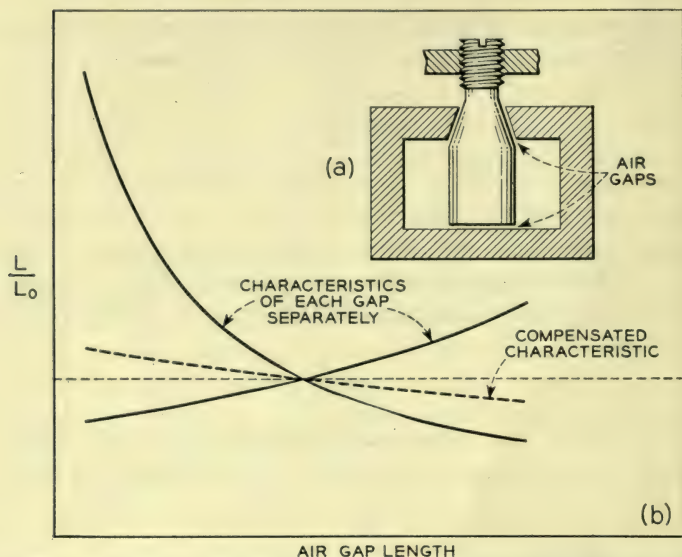


Fig. 12 — Inductance adjustment with partially compensating air gaps.

proximately directly as the ratio of air gap area to magnetic core area. Adjustment by variation of the area of the gap, therefore, is not subject to the nonlinearity that results from manipulating the air gap length. A simple method for providing adjustment by varying the effective area of the air gap is shown in Fig. 13. As one shell is rotated with respect to

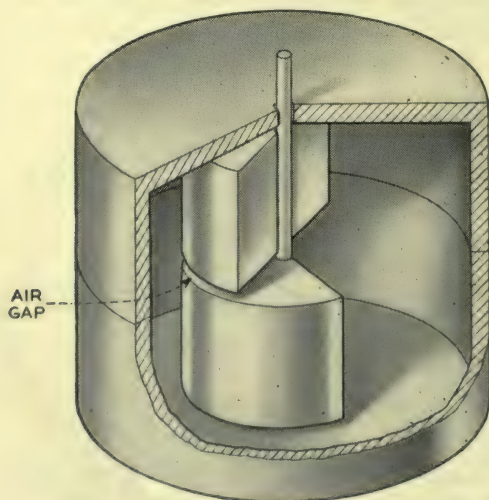


Fig. 13 — Inductance adjustment by variation in air gap area.

the other the area of registration between the semicircular faces of the center posts is reduced or increased. A simple arrangement such as this provides good linearity but has two limitations: (1) The total mechanical motion available for adjustment is only $1/2$ revolution. (2) As the cores are rotated to reduce the air gap area and the inductance, the magnetic flux in the cores tends to concentrate more and more in the vicinity of this reduced area, and may under some conditions give rise to high hysteresis loss.

The method of adjustment used in the 1509 type inductor, Fig. 4, is essentially a variable area method but it is designed in such a way as to overcome these two limitations.⁸ The air gap arrangement, visible in the cutaway view, consists of two gaps in parallel. The main annular gap is fixed and is large enough in area to insure that under no anticipated conditions of operation will the flux concentration be too high. The screw adjustment moves a cylindrical ferrite part into a depression in the center post, as shown. The effective cross section of this adjustable gap is approximately determined by the amount of surface of the cylinder within the depression. The total useful adjusting range corresponds to about six full turns of the adjusting screw.

5. *Adjustment of Inductance by Change in Coupling*

The overall inductance of two coils of equal inductance connected in series is

$$L = 2(1 + k)L_1 \quad (48)$$

where L = series inductance

L_1 = inductance of either coil

k = coupling coefficient

k may have any value between -1 and $+1$, these values corresponding to complete coupling and the windings connected in series opposing and series aiding, respectively. It is practicable to make inductors whose coupling can be continuously adjusted from very high positive values through zero to equally high negative values. This type of design is especially useful where a very wide range of inductance variation is desired. It will be seen from (48) that with couplings of plus and minus 90 per cent, respectively, in the extreme positions a range of 19 to 1 in inductance variation would result. A coil of this type is shown in Fig. 14.

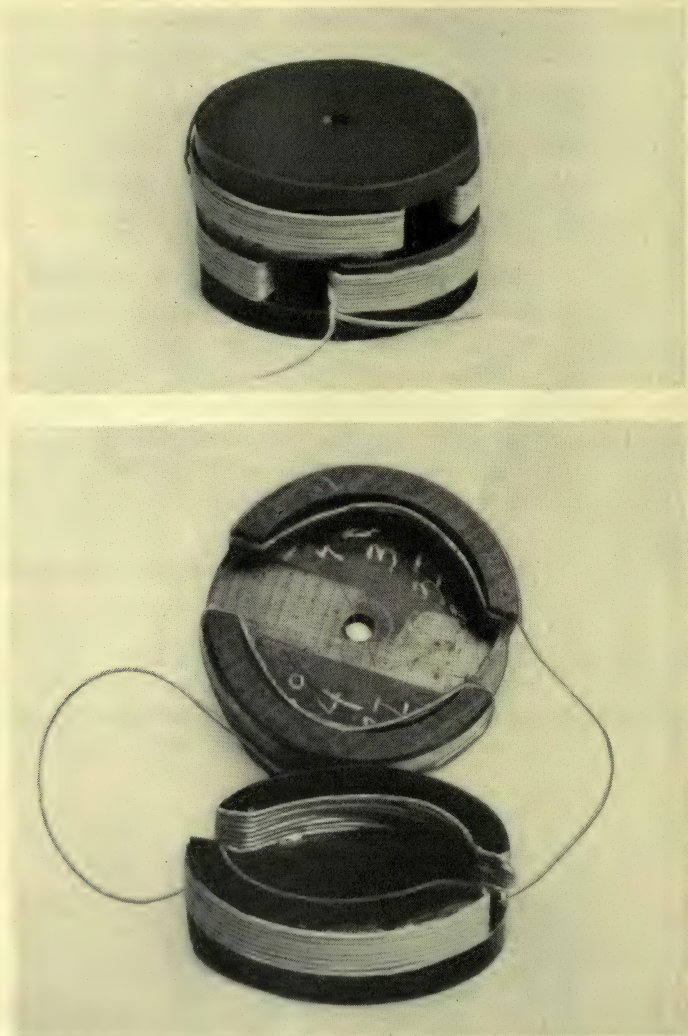


Fig. 14 — Magnetic variometer. Inductance adjustment by variation in mutual inductance.

CONCLUSION

The combined magnetic and mechanical characteristics of ferrite permit the design of inductors superior to those using metallic cores in at least three important respects:

1. Higher values of Q are obtainable than have ever before been practicable. In the range from about 50 to 200 kc, frequently used for telephone carrier, it is very difficult to realize a Q above about 300 in a

metallic core coil. Ferrite coils, on the other hand, have been made having Q 's as high as 1000, and inductors with Q 's between 500 and 600 are available commercially.

2. Formulas have been derived which show how ferrite can be used to realize the best inductor characteristics in the smallest volume. The 1509 type inductor, for instance, is only about $1/3$ as large as the nearest equivalent permalloy core coil, yet its Q is over twice as high.

3. It is physically practicable in ferrite coil designs to include inductance adjustment facilities to meet a wide range of requirements.

It should not be concluded that the day of metallic cored inductors is over. For power applications, especially those involving direct current, present-day ferrites are inferior to silicon iron and permalloy. At voice frequencies there are many applications for which ferrite is, at best, no better than some of the older materials, although in others it has distinct advantages. In higher frequency ranges, however, and especially for low power level applications, the advantages of ferrites are outstanding enough to justify the expectation that they will largely replace the older iron and nickel-iron powders.

ACKNOWLEDGEMENTS

The development of ferrites and their application to inductors has been carried out by several teams in Bell Telephone Laboratories, including J. H. Scaff, F. J. Schnettler and their associates in the metallurgical department, V. E. Legg and C. D. Owens in the magnetic applications group, and S. G. Hale, R. S. Duncan and others in the inductor development area. I don't know who was first to conceive of using "dissipation factor" instead of " Q " to simplify his mathematics, but it was not the author. Equation (24) is derived by inserting $1/D_n$ for Q in an expression originally worked out by P. S. Darnell.

REFERENCES

1. Legg, V. E., Survey of Magnetic Materials and Applications in the Telephone System. Bell Sys. Tech. Jl., **18**, pp. 438-464, July, 1939.
2. Legg, V. E., and F. J. Given, Compressed Powdered Molybdenum Permalloy, Bell Sys. Tech. Jl., **19**, pp. 385-406, July, 1940.
3. Snoek, J. L., Non-Metallic Magnetic Materials for High Frequencies, Philips Tech. Review, **8**, pp. 353-360, December, 1946.
4. Wien, M., Über den Durchgang schneller Wechselströme durch Drahtrollen, Ann. D. Phys. [4], **14**, P. I, 1904.
5. Edwards, P. G., and L. R. Montfort, Type-O Carrier System, Bell Sys. Tech. Jl., **21**, pp. 638-723, July, 1952.
6. Legg, V. E., and C. D. Owens, Patent Application Serial No. 184602, Filing Date — 9/13/50.
7. Hale, S. G., and C. W. Nuttman, Patent Application Serial No. 263564, Filing Date — 12/27/51.
8. Duncan, R. S., and H. A. Stone, Patent Application Serial No. 262248, Filing Date — 12/18/51.

A Throwdown Machine for Telephone Traffic Studies

BY G. R. FROST, WILLIAM KEISTER AND
ALISTAIR E. RITCHIE

(Manuscript received December 3, 1952)

In order to study the traffic-carrying characteristics of the No. 5 crossbar switching system, a machine has been built to simulate the operation of the system. This machine, known as a throwdown machine, is controlled by a team of four operators. Its input is a statistically accurate representation of telephone traffic and its output is a detailed record of the course of each call through the system. This paper discusses the design principles of the throwdown machine, its operation, and the type of results obtained.

INTRODUCTION

Existing analytical methods are inadequate for investigating many statistical problems in which a large number of variables and their interactions must be considered. The problem of evaluating the performance and traffic capacity of a large automatic telephone switching system is one example. Others involve logistics, air and highway traffic control, and certain phases of military and naval strategy. All these require the assimilation of large quantities of data, processing the data according to certain procedures which are often empirical, and producing final information from which performance of the system or the excellence of the procedures can be judged.

These problems fall in the general category of "systems evaluation." The types of systems considered are those that are capable of a large number of variations depending on the nature of the input data, and must be judged on a statistical basis. One method of study might be to operate and observe an actual system. There are a number of objections to this. Operation may be so slow that the accumulation of sufficient data may require excessive time or, as in the case of a telephone switching system, so rapid that it is impractical to make the necessary observations. Operating the system under controlled conditions in these cases may be

too expensive or indeed impossible. Then, too, the system may be proposed only and not yet exist.

One solution is to devise a method of simulating the performance of the actual or proposed system which through the use a suitable time scale will permit the necessary information to be obtained. The simulation may be done entirely on paper by recording each state of the system and modifying this state with each bit of input information according to the system plan, as though a log were being kept of the performance of the actual system.

Since the general problem involves large quantities of input data which are statistical in nature, all possible variations cannot be studied. A sufficient number of typical situations must be tried to obtain statistically reliable results. These methods have been extensively used in telephone traffic studies and are called "throwdown" studies. The name stems from the use of dice in the early study of telephone traffic problems. Each die is designated to represent a particular independent event and the faces of this die are designated according to the probability of the event taking place. By repeatedly "throwing down" a number of such dice and observing the results, the probability of a particular combination of events taking place can be estimated. Other similar methods based on selections from lists of random numbers have been used in telephone traffic studies for a number of years. Recently, mathematicians, using digital computers, have employed similar statistical methods in problems relating to the diffusion of gases, electron ballistics and the solution of certain types of differential equations. They have called this the "Monte Carlo" method.

Various mechanical aids can be used in running a throwdown study. This paper will discuss the techniques of throwdown studies and will describe a semi-automatic throwdown machine which was constructed for studies of the new No. 5 crossbar switching system used in local telephone central offices. A general view of this machine is shown in Fig. 1. It is a system of electrical switching circuits, signal lamps and mechanical devices which simulates a large telephone switching system and its associated subscribers. The machine is controlled by a team of four operators. Artificially generated telephone traffic is processed by this machine in a manner analogous to the action of the actual system. Detailed records are made of the progress of each call and the traffic situations encountered. After a sufficient number of calls have been processed, the recorded information can be analyzed by statistical methods to obtain desired information. The action of the system is simulated in sufficient detail to insure that results are representative of actual

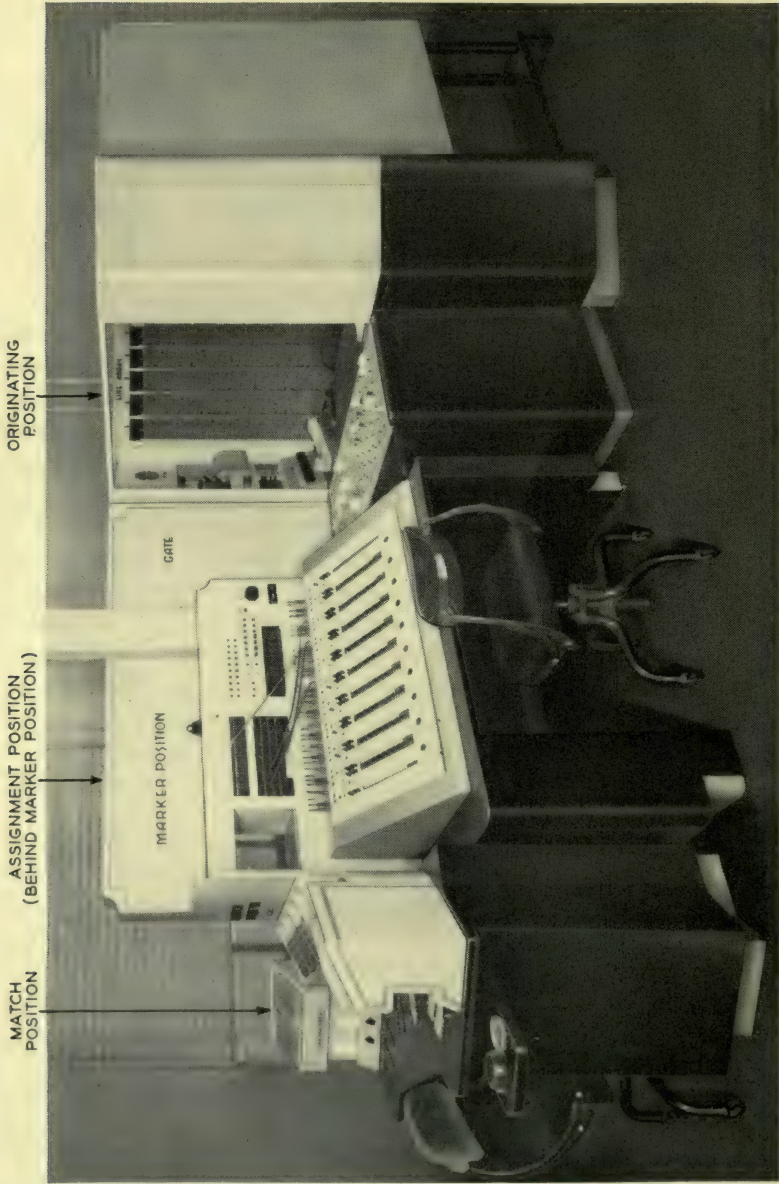


Fig. 1—General view of No. 5 system throwdown machine.

system performance. The level and character of submitted traffic can be varied, and a wide range of system sizes with varying quantities of control circuits can be tested.

Before the throwdown machine was built a "paper" throwdown trial of a small No. 5 crossbar system installation had been conducted by Mr. R. I. Wilkinson. This was run by a team of girls using card files, ledgers and written records, and using dice to make certain random decisions. The machine is basically a mechanization of these early methods to make possible the testing of larger installations in a reasonable time. The methods of generating data for the machine were developed by Mr. Wilkinson and many of the decisions relating to telephone traffic and statistical problems which were encountered in designing the machine were solved in consultation with him.

THE TELEPHONE TRAFFIC PROBLEM

A large automatic telephone switching system of the common control type is not a simple mechanism nor is evaluating its performance and traffic carrying capacity a simple problem. The economy of these systems depends upon the efficient use of relatively small groups of circuits on a time-sharing basis to serve a large number of subscribers. Each group of circuits is specialized to perform certain of the functions necessary in establishing a connection, and circuits from several groups must co-operate to handle every call. A sequence of actions with appropriate alternatives at several stages where busy conditions may be encountered is completely prescribed for every call. However, this sequence is subject to interference due to simultaneous requests to use the same control circuits. Competition is resolved by preference arrangements which cause some requests to be delayed while others are being served. Delays will increase the holding time of circuits with the possibility of causing traffic congestion at other points in the system.

With a number of subscribers originating calls at random, it becomes difficult to predict what the reactions of the system will be at various traffic levels. Although some parts of the problem can be solved by analytical methods employing probability theory, it is doubtful that mathematical means, beyond rough approximations, are available for evaluating the performance of an entire system.

Where systems have been built and placed in operation, the performance can be judged from observations of the working system. There are obvious weaknesses in this procedure. Only by collecting large quantities of information can the performance at various load levels

be determined. Practical systems, as a matter of economy, are not equipped with indicating mechanisms to show performance at all stages. Events take place rapidly and the causes leading up to a particular traffic situation cannot be easily observed. Traffic loads cannot be repeated under controlled conditions. Size of installation and quantities of working equipment can be varied only by small amounts in a working office. To test offices of various sizes requires that these offices be built and installed. Variations in the system operation require actual changes in a working system. This can be done only to a limited extent.

In the case of newly developed systems, estimates of performance can be made on the basis of engineering judgment and experience with older systems of a similar type. This can be followed by a trial installation of a working system of a modest size which will test the system for flaws in design as well as provide information on traffic capacity which can be extrapolated to indicate approximate quantities of equipment for larger installations. At best this is a slow and expensive process and engineering data must be continually revised as experience is gained with larger installations. When the system is to be used extensively in installations of various sizes, methods of evaluating system performance in advance of actual construction are desired. This situation occurred in the development of the No. 5 crossbar system and was the occasion for building the present throwdown machine.

THROWDOWN TECHNIQUES

Since throwdown techniques have played an important part in the development of telephone traffic theory, a brief discussion of the basic principles will be given.

A single throwdown test will indicate the performance of a system under a specific set of conditions. In a typical telephone traffic study, a given traffic load would first be assumed and a simulated system installation to handle this load would be engineered on the basis of the best available information. The test run then will show the performance of the system under these particular conditions, and indicate both the adequacy of the initial engineering procedures and possible improvements. To obtain a proper balance between equipment quantities and traffic load may require several additional runs, varying the equipment quantities, traffic load or both.

The procedure in a throwdown study is to first obtain data representing the traffic to be handled by the system. The traffic data can be generated artificially by the use of random numbers. The method is

based on a knowledge of the statistical behavior of the various factors entering into the composition of real traffic. Random numbers are drawn for each factor. These numbers are assigned values according to frequency distributions obtained analytically or from field observations. The regulating data are combined to produce a description of a sample of traffic which would be encountered under the assumed conditions and then processed by methods which simulate the performance of the actual system.

As a simple example of the throwdown procedures, suppose that it is desired to determine how often on the average an "all trunks busy" condition will occur in a particular group of trunks handling inter-office calls. A certain period of time is first selected and the number of calls expected within this period is determined. Two random numbers are then drawn for each call. One random number specifies the time, within the period, of origination of the call. The other random number, weighted according to an exponention distribution which will be discussed later, gives the holding time of the call.

With the data of call origination times and holding times prepared, the throwdown run can be started. The calls are listed in the order of their originating times. The first call is assigned to the first idle trunk. A record that this trunk is busy is made and the time at which it will become idle determined by adding the assigned holding time to the time of origination. This is also recorded. The call which follows in time of origination is then assigned to the next idle trunk and the process continued for succeeding calls. Before each call is established the release times of all busy trunks are scanned to determine whether any busy trunk should be made idle. In setting up each call idle trunks are chosen from the group in the same order of preference that would be used in the system being simulated.

Thus, the performance of an actual system is reproduced with considerable accuracy and detailed records of this performance can be made. From a study of these records the desired information can be determined. The probability of encountering an "all trunks busy" condition can be found, the average number of trunks busy can be determined, or a frequency distribution chart showing the percentage of the time the number of busy trunks is above any given number can be constructed. If proper records are kept such information as the average number of trunks searched over in locating an idle trunk can be determined. If the trunks were reached through a graded multiple or if they were in sub-groups with a common overflow group, simple extensions of the above procedures would be followed.

This particular problem can be solved by analytical methods and is

presented here only to illustrate the application of throwdown techniques. However, as problems become more complex it becomes increasingly difficult to apply analytical methods. Even in relatively simple systems the interplay of variables may be so involved that existing theoretical methods are entirely inadequate. Various simplifying assumptions must be made and there is often the doubt that some important factor has not been overlooked in formulating the mathematical theory.

The most fruitful use of throwdown methods has been to check results obtained by theory and to obtain data upon which mathematical theories can be based. Throwdown techniques, of course, can be used to obtain direct results, but the functioning of a system will be better understood if there is at least some attempt to develop a theory which explains how various forces act together to produce observed results. Such theories may suggest modifications of the system which will improve its performance.

It can be seen that the planning of a throwdown study requires a thorough knowledge of both the functioning of the system being simulated and the characteristics of the input data being processed by this system. The validity of results will depend upon the faithfulness with which the artificially prepared input data represent real data and the accuracy with which the throwdown routines represent the real system performance. The following two sections of this paper will describe the No. 5 crossbar system and the characteristics of the subscribers using this system. Later sections describe the methods used in the machine for simulating the dynamic performance of an operating system with its subscribers.

THE NUMBER 5 CROSSBAR SWITCHING SYSTEM*

No. 5 crossbar is a marker-controlled system designed primarily for local central office application in the residential sections of large cities and the fringe areas surrounding these cities.

In regions of this type a relatively large proportion of all calls are completed to subscribers within the same office. Since the surrounding offices to which connection must be made are likely to be of widely diversified types, the system is designed to interconnect with any existing type of central office. No. 5 is also capable of serving isolated centers from about 3,000 lines up, and multioffice areas including the largest

* F. A. Korn and James S. Ferguson, *Trans. A.I.E.E.*, **69**, Part 1, pp. 244-254, 1950.

metropolitan business exchanges. Although the system includes facilities for toll and tandem switching, these were not included in the throwdown studies and will not be discussed.

The Switching Network

The No. 5 system is built around an interconnecting network consisting of two types of switching frames utilizing crossbar switches and known as line link frames and trunk link frames. This is illustrated in block diagram form on Fig. 2. Each frame is double-ended and provides means for connecting any point on one side of the frame to any point on the other side. The connecting paths are known as links.

All subscriber lines in the office connect to one side of the line link frames, each of which can serve, roughly, 300 to 500 lines; and all trunks connect to one side of the trunk link frames, each of which has 160 trunk appearances. The other sides of line and trunk link frames are connected together in such a manner that each line link frame has access to all trunk link frames over several paths.

These interconnecting paths are known as junctors. The basic maximum number of line link and trunk link frames is 20 and 10 respectively, and this is the size embodied in the throwdown machine. However, in practice, multiplying arrangements can be employed to double the number of frames to give greater subscriber and traffic capacity.

With the system described above, any subscriber line can be connected to any trunk over one of several paths, each consisting of two links and a junctor and known as a channel. On connections to outgoing or incoming trunks, a single channel is required; on connections through intraoffice trunks (connection between two local subscribers), two channels, one from each end of the trunk, are required. The method of combining links and junctors to form channels will be described later.

Dial Registration—Originating Registers

The circuits that receive and store the dialed signals from the subscriber are known as originating registers. These circuits, in quantity as determined by desired quality or grade of service, are distributed over the trunk link frames as equally as possible. A connection between subscriber and register is set up through the switching network just as between subscriber and trunk. The registers call in control equipment for setting up the subscriber's connection when dialing is completed.

Common Control Circuits—Markers

The switching of all connections in the office is performed by a group of common control circuits known as markers, any one of which may be utilized on a particular call. The principal functions of the marker are (1) to determine or receive the specific location of a calling circuit; (2) to translate input signals into the specific location of a called circuit or group of circuits; (3) to test for availability and seize a called circuit or one of a group of circuits; (4) to locate, test and seize a switching path between calling and called circuits; (5) to set up the connection; and (6) to take alternative action in case of trouble or busy conditions. A marker performs these functions in a very short period of time so that a few circuits can handle the requirements of an office. In the original design of No. 5 crossbar, a single type of marker handled all connections. This was the arrangement specifically handled by the throwdown machine. Later design has introduced three types of markers; dial tone, completing, and combined.

As an example of the function of the control circuits, when a subscriber originates a call, a connection is automatically established from the subscriber line circuit on a line link frame via a marker connector to an available marker. The marker identifies the location of the line and establishes the fact that it is a new call requiring a register. It tests all registers and trunk link frames and chooses an idle frame with idle registers. The marker then gains access to the correct line link and trunk link frames via the frame connectors, chooses an idle register, tests all usable channels, picks a particular channel and operates the crossbar switch magnets to close the connection between line and register. After storing the line location in the register for later use, the marker disconnects itself.

When the subscriber completes dialing, the register connects itself to an idle marker via the marker connector. It transfers to the marker the location of the originating line and the called number. If the call is local to the office, the marker determines the location of the called line from the number group circuit (a translating device) and tests and chooses an intraoffice trunk. The marker then gains access to the link frames through the frame connector, tests the called line for busy, picks a channel, and establishes the connection, thereafter removing itself and the register from the connection.

During the course of the foregoing events, a call may encounter various delays beyond the minimum circuit operating time in setting up the connection. Delay may be caused by meeting a temporary busy condi-

tion of markers, registers or the particular number group link frames required. Busy condition of lines or trunks may result in rerouting the call. In general, the grade of service is measured by the delay in connecting a subscriber to a register (dial tone delay) and the probability of not finding an idle trunk or channel.

Intraoffice Trunks

The intraoffice trunks, used on locally originated and completed subscriber connections, include the supervisory, charging and ringing functions. The trunk is held on a connection for the duration of the call in distinction to markers and registers which have short holding times.

Outgoing Trunks and Senders

Calls to other offices are completed via outgoing trunks which incorporate the supervisory and charging functions. In order to transmit information to the distant office, an outgoing sender is connected to the trunk for a short period of time by means of an outgoing sender link. In establishing the call, the marker first connects itself to the sender via the sender connector in order to register in the sender the called number, and then sets up the trunk-sender linkage. When the sender, which may be one of several varieties depending upon the type of signaling required by the distant office, has transmitted the number to the distant office, it drops off the connection.

Incoming Trunks and Registers

Calls from a distant office are completed through an incoming trunk, which includes supervisory and ringing functions. In order to receive signals from the distant office, the trunk connects itself to an incoming register by means of an incoming register link. When the register, which again may be one of several varieties, has received the called number, it obtains a marker through the marker connector for setting up the connection. When its functions have been accomplished, the register disconnects.

Tone Trunks

This group of trunks includes those to which calls are routed when line busy or all trunk or channel busy conditions are met. Subscriber error or excessive delay in dialing also result in routing to these trunks. The tone trunks return distinctive tone signals to the subscriber. If

the marker is unable to set up a connection to a tone trunk, the originating register is able to return the tone signal.

Number Groups

Each number group is a translating circuit which provides terminating information for a consecutive block of 1000 directory numbers. When a marker transmits the called line number to a number group, it receives back the line location on a line link frame, the type of ringing required, whether or not the line is in a PBX group, and certain minor information. The number group also includes facilities for selecting an idle line in a PBX group. As with link frames, only one marker can connect to a number group at a time, but markers can connect to different number groups simultaneously without interference.

Connectors

The marker connector, providing access from line link frames and register to markers, include circuits which assign preference of all line link frames and registers for specific markers. This helps distribute marker usage.

In addition, when several line link frames and registers are competing simultaneously for busy markers, a gating arrangement allocates the order of service to reduce excessive delays. Although shown as one block on Fig. 2, the marker connector circuits are provided one per line link frame and one per sub-group of ten or less registers.

The frame connector, which provides access from markers to link frames, includes preference and lockout features since only one of several competing markers can connect to a given frame at one time. Although shown as one circuit on Fig. 2, the frame connectors are actually distributed over the frames.

Outgoing senders of a particular type are divided into two groups and the sender connector permits connection from two markers, each to one sender in each group simultaneously. Preference and lockout features obtain.

The number group connectors, provided one per number group, are similar to frame connectors and give access from markers to number group.

Handling a Typical Call

An understanding of the operational intricacies of a telephone switching system cannot be gained by a discussion of components and can only

be developed by a study of the course of events in setting up calls through the system. As has been intimated earlier, the establishment of calls is largely a matter of marker operation. In order to illustrate the marker functions, Figs. 3 and 4 show two charts indicating the order of events in establishing a connection, first, between a calling or originating subscriber line and a register, and second, after dialing, between two subscriber lines within the same office. These two types of connection are known as a dial tone call and an intraoffice call, respectively.

Figs. 3 and 4 are drawn as sequence charts with time flowing downward. There is no attempt to maintain an accurate time scale; the x marks on the vertical line merely represent the relative order in which important control functions take place. In actuality, of course, the time between x marks is known with fair precision. Brief descriptions of the control functions are listed to the left of the vertical lines. The call illustrated is presumed to encounter no difficulties in completion. However, points at which blocking might occur are marked with an asterisk to the right of the lines. If any of the difficulties noted were to develop, the marker would have to take alternative action which will be illustrated later. Also shown to the right of the lines are potential points of delay, where a call may have to wait until a connector, a marker, or a desired frame becomes idle. It must be remembered that during moderate or heavy traffic, several or all of the markers are working simultaneously and tending to interfere with each other.

In a well-balanced and soundly engineered central office, the aggregation of parts are nicely adjusted to give on the average no more than certain preassigned values of delay and blocking at some average busy hour traffic level chosen as a base. A typical example of permissible delay is no more than 1 per cent of calls having a dial tone delay greater than three seconds. When traffic is heavier than the engineering base, the percentages of delay or blocking will increase.

A summation of all the possible alternative sequences which a marker may have to take when trunk busy, line busy and channel busy conditions are encountered becomes extremely complex. Although no attempt will be made to discuss this in detail, a chart showing the operational variations of a marker on an intraoffice call is presented on Fig. 5. Even this figure does not include all possible variations since, for example, the contingency of all tone trunks being busy is not shown on the diagram. This chart, similar in form to Fig. 4, will later be found useful in discussing the throwdown machine. In order to simplify the presentation, some of the control events are combined in time with the frame seizure which precedes the event. The normal course of a call

SUBSCRIBER LINES
MAX - 500 PER
LINE LINK FRAME
TOTAL MAX - 10,000

LINE LINK FRAMES
MAX - 20

JUNCTORS

TRUNK LINK FR
MAX - 10

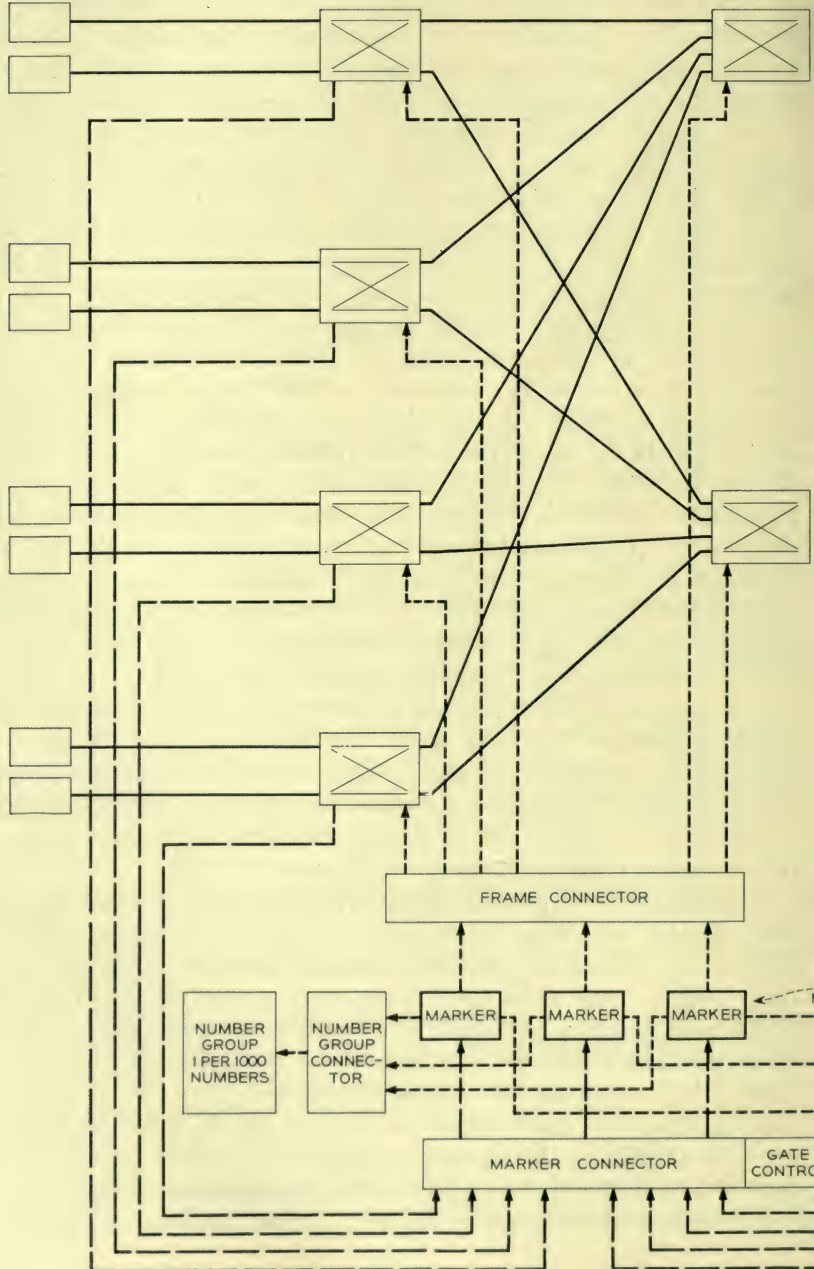
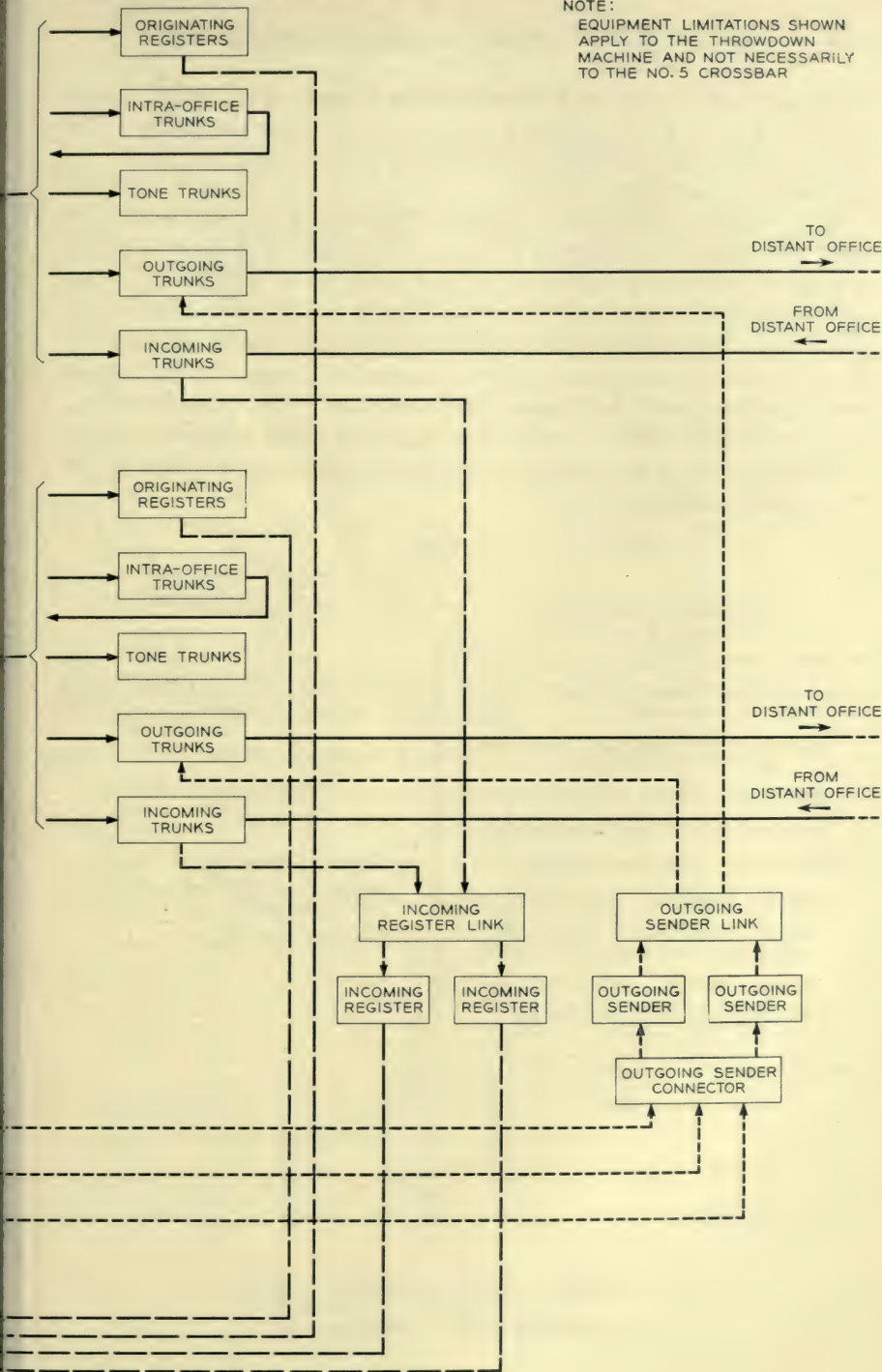


Fig. 2—Principal con

TRUNKS-ORIGINATING REGISTERS
 -160 APPEARANCES PER TRUNK LINK FRAME
 (A-OFFICE TRUNK REQUIRES TWO APPEARANCES)

NOTE:
 EQUIPMENT LIMITATIONS SHOWN
 APPLY TO THE THROWDOWN
 MACHINE AND NOT NECESSARILY
 TO THE NO. 5 CROSSBAR

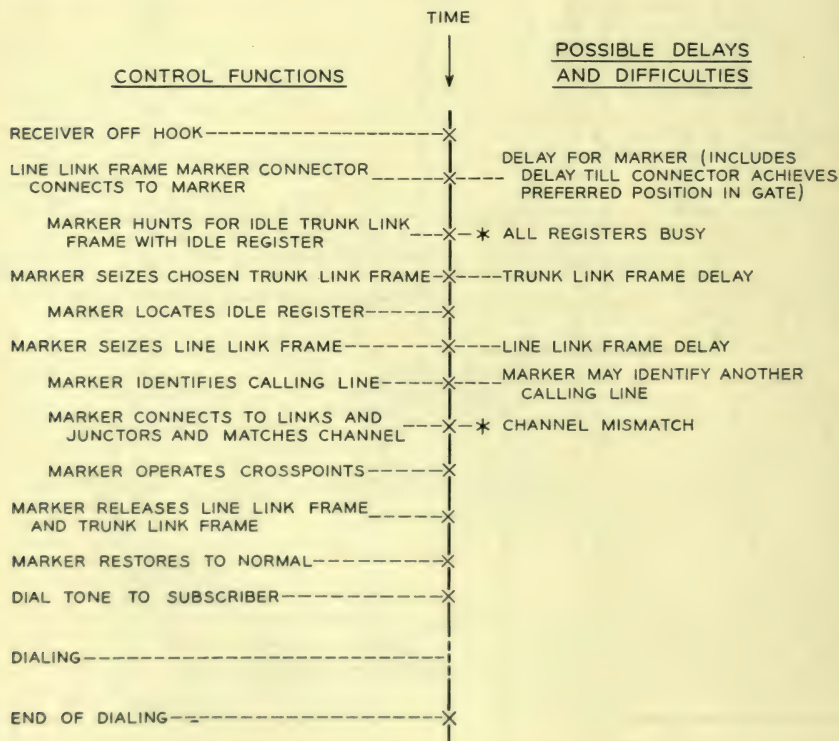


5 crossbar office.

is in a vertical direction and it is only when a busy condition or a special function such as PBX hunting is encountered that the call shifts to the right or left. On a particular call, the only control functions performed are those marked with an x.

In the situation illustrated, the size of the office is assumed to be such that two groups of intraoffice trunks are provided. The marker makes a more or less random choice of the first group to be tested, and, if this group proves to be all busy, will automatically test the second group.

This brief description of the Number 5 Crossbar System is not intended to be exhaustive. It discusses only the important features of the system and those which will help make the description of the throwdown machine more intelligible. Certain more involved items will be discussed in greater detail in later sections concerned with the functioning of the throwdown machine.

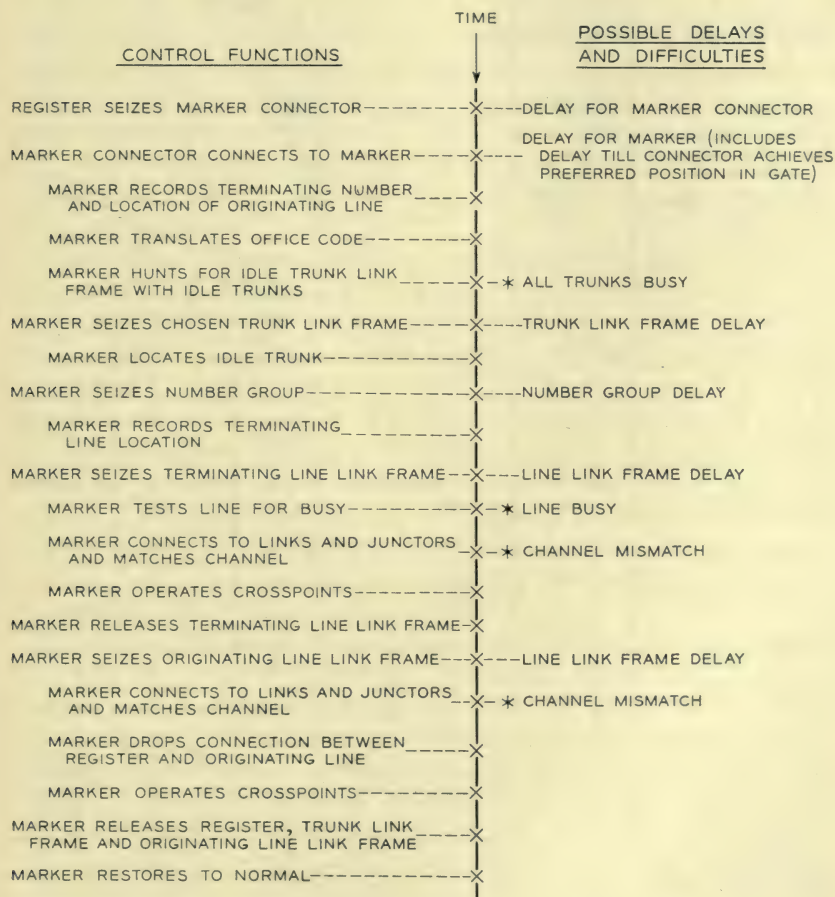


* MARKER MUST TAKE ALTERNATIVE ACTION IF IT ENCOUNTERS THIS CONDITION

Fig. 3—Establishing a dial tone call.

CHARACTERISTICS OF SUBSCRIBERS AFFECTING DATA

Since the purpose of the throwdown machine is to evaluate performance and traffic capacity of a simulated switching system under conditions met in service, subscriber data fed into the machine must represent, as nearly as possible, the characteristic actions of real subscribers. Fortunately, telephone subscriber characteristics can be studied in working switching systems which are similar to the one under throwdown evaluation. Little error is introduced by such subscriber data



* MARKER MUST TAKE ALTERNATIVE ACTION
IF IT ENCOUNTERS THIS CONDITION

Fig. 4—Establishing an intraoffice call.

inasmuch as subscriber behavior is very slightly influenced by the type of system serving their telephones.

Subscribers, although they are individuals, exhibit many "group characteristics" dictated not by the requirements of telephone communication but by their mode of life. This fact allows statistical treatment of many observed action distributions without introduction of significant error. However, these group actions also present problems of congestion in telephone plant which require detailed throwdown study for solution.

As an example of group characteristic, subscribers do not originate a steady barrage of calls over the twenty-four hours of the day. During mid-morning and mid-afternoon hours traffic is built to a peak value, whereas during certain of the remaining hours it is reduced to a minimum. In some residential areas peak traffic may also occur during the early evening. Throwdown evaluations of simulated switching systems, however are primarily concerned with the busy hour, the hour in which the greatest number of calls are originated, regardless of its actual time of day occurrence.

Useful datum obtained from busy hour field observations is the calling rate per subscriber (calls per hour) which can be used to set up traffic load conditions on the simulated switching system. The calling rate characteristics can be measured as average calls per hour placed by subscribers in a number of group classifications. An example used in a particular throwdown study of simulated system response to a given traffic load is given in Table I. The values given in this table represent average day to day calling rates. Weather conditions, pre-holiday periods or special events have been found to raise substantially the average calling rate in affected classifications. Values adjusted for these conditions are useful in projecting percentage of overload that can be offered to systems engineered for average daily loads.

Subscribers, however, in originating calls, act independently within their classified group in maintaining the average calling rate. Originating times of calls, therefore, occur at random within the hour. Throwdown input data representing subscriber originating time behavior are produced by assigning to each call, of the total within the studied hour, a six digit number from a list of random numbers. If the hour is divided into one million parts the assigned random number determines the millionth part of the hour in which each call will originate.

Observations made in the field have shown that subscribers, upon receiving dial tone, do not always follow through to dial a full code. Among possible causes are failure to hang up after completion of a call, answering the wrong telephone where two or more are adjacent, dialing

TABLE I

Group Classification	Average Calling Rate Per Subscriber During Busy Hour
Heavy demand individual line subscribers (such as doctor and professional services).....	5.0
Medium demand individual line subscribers (such as small business and some residential subscribers)....	0.85
Light demand individual line subscribers (mainly residential subscribers).....	0.02
P.B.X. line subscribers (such as large businesses, hotels, railroads, etc.).....	3.0

before dial tone, and forgetting the number. Such actions produce waste usage of equipment within the switching system, and their study is pertinent to producing throwdown data. To simplify this study, all subscriber actions involving the alerting of central office equipment are designated "subscriber starts" and divided into four categories as given in Table II.

Since "no dials" and "partial" dials are largely due to subscriber errors in properly originating calls, many of these calls will be originated upon discovery of the error. False starts, on the other hand, are attributed to accidental origination with no intent to place a call. A flow chart illustrating these actions is shown in Fig. 6. The importance of this subscriber behavior is indicated by the percentage of waste usage calls (FS, ND, and PD) to ultimate good calls. Pen recorder tapes taken at particular central offices showed waste usage calls at 30% and good calls at 80.5% of ultimate good calls. For these specific cases false starts represented 7.5%, partial dials 7.5%, and no dials 15% of ultimate good calls.

TABLE II

Category	Description
Good calls	Calls on which the subscriber waits for dial tone and then dials the required full code
False starts	Calls, on which a sender or register is seized, but which are abandoned in less than two seconds without dialing
No dials	Calls lasting longer than false starts but on which no dialing occurs. These calls may exceed a certain length "time-out" period and be given a distinctive tone
Partial dials	Calls on which less than the required full code is dialed. These calls may be held beyond a certain length "time-out" period and be given a tone

It was also found that approximately 90% of the partial dials and the abandoned no dials were reoriginated. These percentages are quoted only to illustrate subscriber behavior under certain conditions of central office load at a particular office. Type of service, load conditions on the central office, and location can effect these percentages. For a more detailed analysis, see "Dialing Habits of Telephone Customers."*

Fig. 6 illustrates only the group behavior of subscribers. Individually, the subscribers will hold equipment on abandoned no dials, abandoned partial dials, and false starts for varying amounts of time. These varying individual holding times can be quantized into several average values which are equally likely to occur, or may be averaged to one value as shown on Fig. 6 depending upon the throwdown study requirements. The holding times on calls receiving tone are usually assumed to cease a few seconds after tone is received.

Subscribers, as individuals placing ultimate good calls, spend varying amounts of time, after receipt of dial tone, before start of dialing and

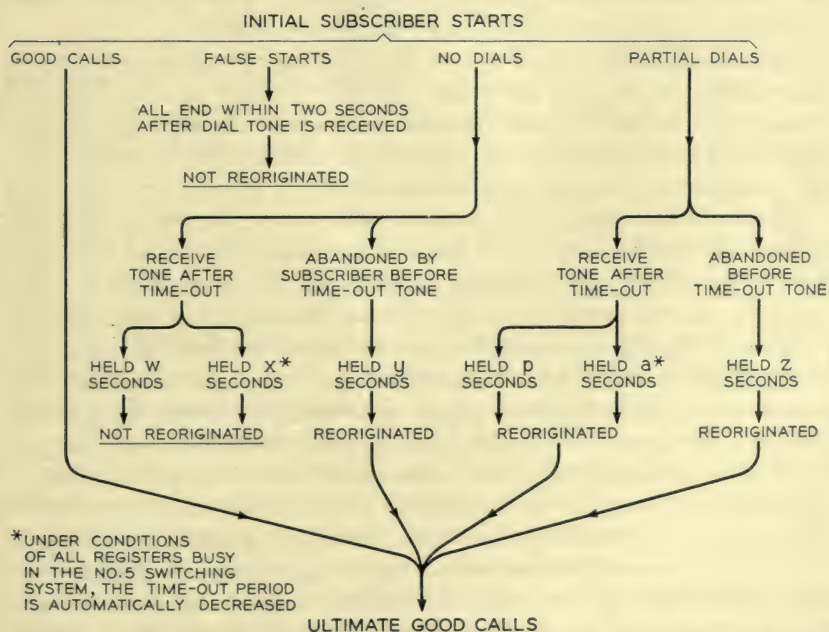


Fig. 6—Simplified characteristic action of subscribers in converting initial subscriber starts to ultimate good calls.

* Charles Clos and Roger I. Wilkinson, "Dialing Habits of Telephone Customers," Bell System Tech. J., 31, pp. 32-67, Jan. 1952.

in dialing a full code. This behavior affects the holding time of registers receiving the dialed digits and must be considered in throwdown studies. Data collected on combined waiting and dialing time characteristics show a frequency and time distribution that can readily be quantized into a number of values, each equally likely to occur. When the number of values for a particular throwdown study are determined, each quantized dialing time is represented by a number. Each ultimate good call is then assigned a number from a random list of the representative numbers to establish the dialing time of the call.

Ultimate good calls will develop one of three terminating conditions attributable to subscriber behavior: 1) DA, called subscriber does not answer, 2) busy tone because of called subscriber line busy, or 3) answer by called subscriber. It is assumed from analysis of "don't answer" studies that, for certain throwdown evaluations of the switching system, approximately 10% of the ultimate good calls meet the DA condition. The number of busy tone terminations, of course, will develop during the throwdown study as a result of the average originating and terminating calling rate per subscriber served by the system.

Most subscribers, upon encountering a line busy condition make subsequent attempts to reach the called line. The number and frequency of attempts made depend upon the individual characteristics. A detailed analysis of this characteristic, suitable for use in throwdown studies, has appeared in a paper by Charles Clos.*

When calls are answered by called subscribers, the connections will be held for varying amounts of time. It has been determined from field observations that the frequency distribution of these holding times is closely approximated by an exponential distribution. For throwdown purposes a simplifying assumption can be made that holding time is not a continuous variable but is quantized so that a particular holding time will have one of several values. To determine these values an exponential distribution having the proper average is plotted as shown in Fig. 7. The area under the curve is then divided into the required number of equal parts (ten, for this example). A central value of holding time is determined to represent each subarea. Ten holding times are thus produced which are weighted according to the exponential distribution and which are equally likely to occur. These holding times can be designated 0 to 9 and assigned to the calls by choosing one-digit numbers at random for each call.

* Charles Clos, "An Aspect of the Dialing Behavior of Subscribers and Its Effect on the Trunk Plant," Bell System Tech. J., 27, pp. 424-445, July, 1948.

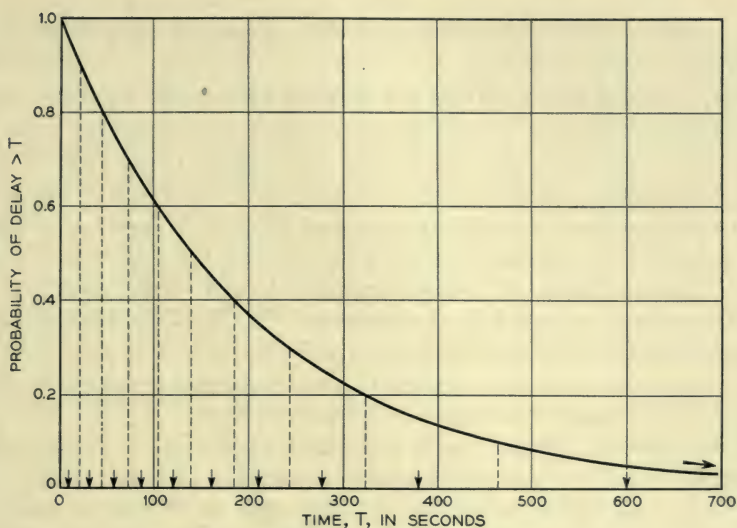


Fig. 7—Distribution of holding times.

GENERAL PLAN OF THE MACHINE

The broad plan of the throwdown involves a division of work between the team of operators and the circuits of the machine. The circuits keep track of system events, resolving complex sequences of actions concurrently taking place. Their chief purpose is to present signals to the four operators so that they will be able to perform the right actions at the proper time and thus dovetail together the events for a large number of calls in progress.

The operators keep manual records of the busy-idle states of items which occur in large quantities, such as lines, trunks and links. They also perform the searching operations necessary to locate these items when they are required to be made busy or idle.

In general the circuits signal the operators to perform actions; the operators in turn signal the circuits that the action is completed or some appropriate alternative taken. The circuits then determine the next action and present corresponding signals to the operators. Thus the circuits largely control the sequencing of events. However, in some cases where the sequencing would require extensive circuitry, the operators assist in determining sequence. For example calls returning to the control circuits after a subscriber completes dialing are interleaved with new calls according to written records maintained by the operators. Releasing

connections are also placed in proper time sequence by the operators according to written records.

The actions of the operators are checked electrically in many cases. An improper setting of certain switches, which is inconsistent with the state of the system at the time will in some cases give an alarm and in others, block the progress of the machine until the error is corrected. Certain actions performed by the operators involve the insertion of plugs into jacks. Where alternative actions are possible in response to a given signal, separate groups of jacks correspond to the several alternatives. Insertion of a plug into a particular group will signal the circuits as to which alternative is taken. The circuits make one of several keys effective and the operator must then depress this key (corresponding to the action she has taken) before the circuits will advance.

Where several operators must cooperate to perform a given action, signals are passed between the operators by means of keys and all operators must respond before the circuits can advance. Wherever possible, overlap operation is employed. Signals are presented to several operators simultaneously and each operator starts the indicated action, signaling the circuits when it is completed. When the signal from the last operator is received, the circuit advances. In some cases an operator is allowed to start a particular action before the stage has been reached at which this action is required. For example the information necessary to choose links and determine a suitable path through the interconnecting network of the No. 5 crossbar system is available before it is necessary to establish the connection. Since this search is time consuming, the operator is allowed to start as soon as the information is available. At the proper time she is given a signal to complete the record of this connection, or if the call has been blocked before reaching this stage, she is given a "back out" signal instructing her to restore her records to their previous condition. Since this is a rare condition occurring less than one time in 100 tries, little useless work is done and the overall action is speeded up.

A block diagram showing the relations between the operators and the various major components of the machine is shown in Fig. 8 The CLOCK controls the flow of time in the machine and gives an indication of simulated present time in the traffic run being tested. The TIME DETECTORS, of which three are used, provide a means for the operators to indicate a future time at which some action must be taken and be signaled when this time arrives. The block labeled CONTROL CIRCUITS, which present action signals to the operators, represents the main body of circuits which maintain a current record of the states of the various

complex units of the system, such as markers. The GATE provides a means for operators to feed traffic into the machine and determines the order in which each working item of traffic is taken up by the control circuits. The RANDOM CIRCUIT is an electronic "roulette wheel" which permits the operators to make random choices in disposing of certain traffic items. It provides random selections varying from one out of three to one out of ten.

The individual record of each call is made on a card of the form shown in Fig. 9 which is called a "call slip." These are of various types identified by distinctive designations and colors for the several varieties of calls which may occur. The basic types are: dial tone, intra-office, incoming, and outgoing.

As the call progresses through its various phases the slip is passed from one operator to another, each operator retaining the slip while it is in the phase under her control and recording on it in designated spaces the nature and time of the events taking place. When the call is completed, the slip carries a complete record of the call including the designation number and time of seizure of the various circuit units used in establishing the connection and the nature and duration of any delays encountered.

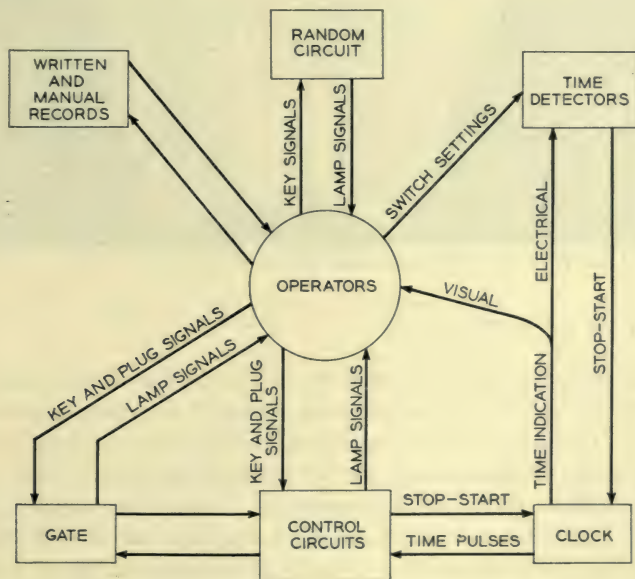


Fig. 8—Communication between the machine components and the operators.



Fig. 9—Call slips on which records are made in the throwdown machine.

DESCRIPTION OF POSITIONS

The throwdown machine consists of five positions as shown in the plan view of Fig. 10. A photograph showing all but one position appears on Fig. 1. This division into positions results from the requirements of simulating the components of the No. 5 crossbar system and equalizing the work load on the attending operators. The arrangement of positions, as shown, provides a continuous clockwise flow of records and other items that must be passed from operator to operator. The five positions are known as: originating position, gate position, marker position, match position, and assignment position. Four operators attend the five posi-

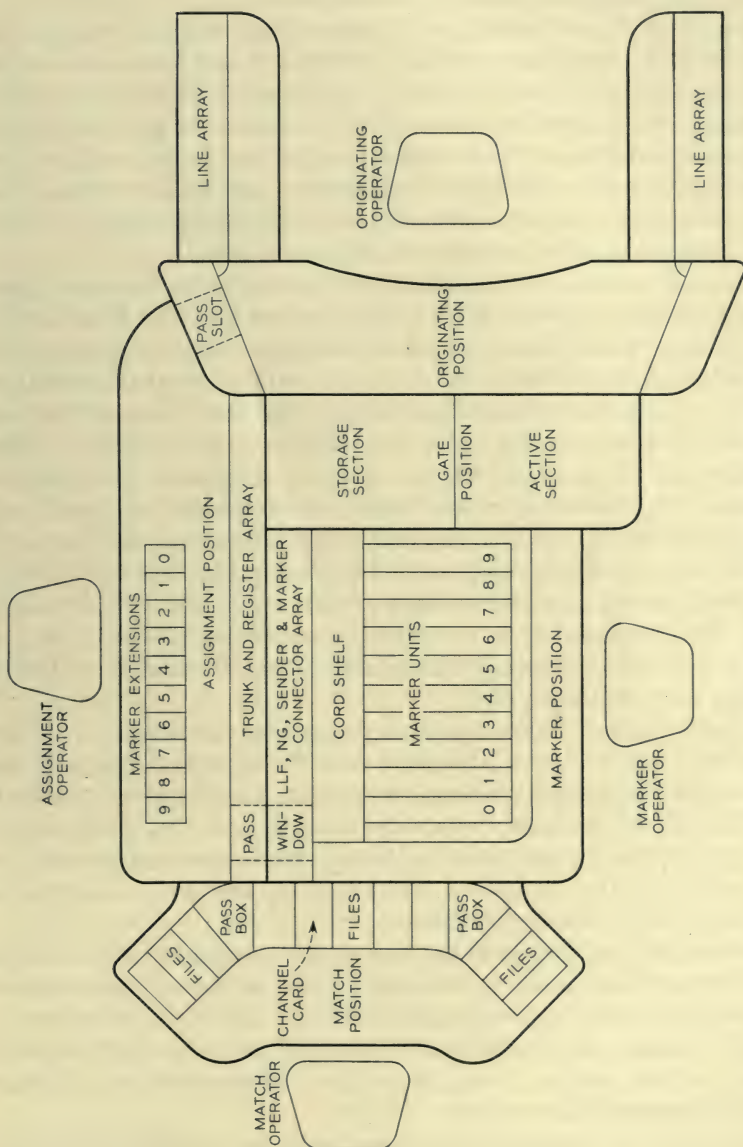


Fig. 10—Plan view of throwdown positions.

tions, the gate position being jointly served by the operators stationed at the originating and marker positions. Each operator is given the same name as the position at which she is stationed.

The general plan of traffic flow is such that all calls are originated or restarted at the originating position. For this reason, the major equipment items of this position include: the subscriber line array, which represents by pegs all of the subscriber lines associated with the switching system under throwdown evaluation; the trunks incoming to this system; and certain of the system components associated with incoming calls. The call slips, sorted in sequence as to originating times, are stored at the originating position for use in originating calls.

Means are provided for presetting the times at which the next originated or restarted calls will enter the throwdown machine. When actual time coincides with a time thus preset, the system stops and signals the originating operator to serve the waiting item. The originating position does not provide for actually entering an item into the machine, but indicates the time of entry, stops the machine, and supplies the items to be entered. For example, the operator, when signalled to originate a dial tone call in accordance with information furnished on the call slip, selects, from the line array, a peg representing the subscriber line, associates it with the call slip, and passes the two items to the gate for entry.

Since the means for determining the busy or idle condition of a subscriber line are provided by the presence or absence of that line's peg in the array, the originating position also enters into the operation on incoming and intraoffice calls.

The gate position, which simulates the marker connectors of the No. 5 crossbar system, serves as the entry point for all calls. The originating operator inserts call slips and pegs into the gate position from one side to start the call. The marker operator removes them at the other side for processing in the marker position. Relay circuits associated with the gate position control the flow of traffic through the gate in accordance with actual No. 5 crossbar operation.

The marker position provides means for associating the call slips with the individual simulated markers of the switching system. Since these markers control processing of the calls, the principle records of the calls' progress are obtainable through this association. The records are kept as time entries on the call slip and marked adjacent to action lamps determining these entries.

At the top of each marker unit in the marker position are cords which provide access to the switching system components under control of the marker. These components are line link frames, number groups, sender subgroups, and marker connectors. As the call progresses, the

marker operator will connect and disconnect the marker with these components as directed by the action lamps provided.

The principle purpose of the assignment position is to provide equipment for test and choice of a trunk on each call. A trunk jack array which includes all trunks and registers of the switching system therefore appears on the face of this position. Since, in the actual No. 5 crossbar system, testing and choosing of trunks are performed by the markers, an extension of each marker also appears in the assignment position. The assignment operator, when required, makes a visual test for idle trunks to the proper destination. She then chooses and associates one of these idle trunks with the active marker by means of a marker cord which is plugged into the trunk jack.

In addition to this principal function of trunk choice, the assignment position is equipped to determine the disposition of calls when the marker has finished setting them up. Means for ascertaining conversation time, dialing time and other types of holding times is provided.

The positions so far described have simulated only the two ends of a connection, the subscriber line and the trunk or register. To complete the connection a channel must be set up between these ends. In the actual switching system the marker matches a line link, a junctor, and a trunk link to form the channel. The match position is provided to simulate this action. The match operator, through visual inspection of a set of channel cards, tests and makes busy the channels for each connection. Information as to the originating subscriber line reaches her through the pass box in the form of the upper portion of the call slip, Fig. 9. Information as to the trunk or register choice is passed verbally by the assignment operator. Since this operation is a function of the marker, marker extensions appear at the match position. These extensions are provided with action lamps to guide the match operator and with keys to inform the simulated marker circuits of the results of the match.

A pass window is cut between the marker position and the assignment position for passage of the call slip and originating peg at the time of marker release. Similarly, a pass box is provided for passing the upper portion of the call slip (match ticket) from the match operator to the assignment operator. The assignment operator is charged with the disposition of these items.

THE TIME SYSTEM

The timing system of the throwdown machine is based on a start-stop system of time pulses. Pulses generated by the clock, Fig. 11, drive indicators which display time and also drive circuit elements which

count time units to cause events to occur in the proper sequence. When these time counters reach a stage where some action is to be performed by the operators, "stop-time" signals are produced. These signals lock in and cause the clock to halt. Simultaneously, action signals are displayed to the operators. When each operator completes the indicated action she depresses an appropriate key at her position which extinguishes the action signals and removes the stop-time signal. When the last operator has responded, all stop-time signals are removed, the time lock is opened and the clock again advances. Thus the clock, in effect, takes time-out while the operators perform the various manual searching and recording actions necessary to simulate and tabulate the performance of the crossbar system.

In the throwdown machine, each clock pulse represents one millionth of an hour or 3.6 milliseconds. This quantization of time is based on two considerations. The first is that it is convenient to represent a particular time during an hour by a six-digit decimal number. The second is that the time represented by one time unit (3.6 milliseconds) is well under the average acting time of the relays and switches used in the system being simulated. Thus the events taking place in the actual system can be reproduced in sufficient detail. Some events of longer time duration

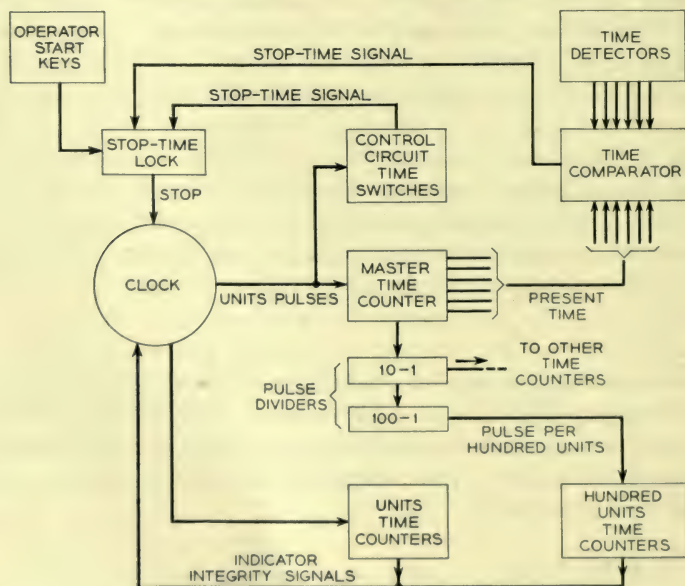


Fig. 11—Block diagram of the time system.

are timed in less detail. For this purpose the clock is also arranged to deliver pulses at one tenth and at one hundredth of the basic pulse rate.

The clock, circuit wise, is a form of free running relay pulse generator. It consists of a series of relays in which the first, in a released condition, causes the remainder of the series to operate in sequence. When the last relay of the series operates, it causes the first relay to operate. The remainder then release in sequence. When the last relay releases, it causes the first relay to release. This cycle, if not interfered with, is repeated continuously to produce pulses representing units of time. Time, thus, can be stopped by the simple expedient of allowing the stop time signals to hold or "lock" the first relay in the operated state.

Time is visually indicated in units and hundreds of units at the operators' positions by a group of telephone type message registers termed the *time counters*. Certain of the registers indicate present machine time for action recording purposes. Others are set at a specified number of units ahead of present time to indicate future times at which held items will be released or re-entered into the system:

To drive these counters and to safeguard their integrity, the counters are substituted for the last relay in the clock pulser relay series. The operating windings of all *units counters* are connected in parallel. Contacts, which make on each counter when the individual counter is advanced, are all connected in a series circuit to form the last relay contact. Failure of any units indicator to advance will, therefore, interrupt the pulse generator cycle and stop time until the trouble is cleared.

The integrity of the hundred units time counters is guarded in the same manner. On each hundred pulse when these counters are advanced, their windings and contacts also form a part of the pulser circuit.

The basic pulse repetition rate of the clock is approximately four pulses per second, being determined by the acting time of the counters and the various circuit elements which the clock pulses must drive. Since each pulse represents 0.0036 seconds, the ratio of basic machine time to real time is in the order of 70 to 1.

The clock pulses, Fig. 11, are counted by two types of time switches. One type, the control circuit time switch, is associated directly with the component control circuits of the throwdown machine. These time switches are not continuously driven but are automatically connected to the clock as required to time events in the progress of a call. The control circuit time switches, as discussed in more detail later, are returned to zero after each usage in preparation for timing the next event.

The second type of time switch, designated the *master time counter*,

is continuously driven by the clock. This counter records simulated present time throughout the entire running of a throwdown test.

The master counter does not, in itself, generate stop time or action signals. Its primary purpose is to furnish correct present time to the time comparator circuit where such signals are generated. It also controls a pulse divider which furnishes pulses at one tenth and at one hundredth the basic pulse rate.

Since failure of the master time counter to advance on each pulse would introduce present time errors which are cumulative over the throwdown run, the integrity of this circuit is rigidly guarded. Checking circuits verify that each basic clock pulse advances the master time counter. The checking circuits, not shown on Fig. 11, upon detecting a failure to advance, produce a stop-time signal and an alarm signal which can be released only after the counter is brought to correct time. The master time counter consists of pulse driven rotary switches arranged so that each switch represents one decimal digit of time. To count one hour of simulated time, as provided in the master time counter, six switches are necessary. These switches record 000000 to 999999 units of 0.0036 seconds.

Means are provided for presetting such items as subscriber originating times, incoming call originating times, and hold release times by the *time detectors*. When a time so set coincides with simulated time, the clock is stopped by a stop-time signal and an action signal indicates that a call is to be originated or a held item released.

The time detectors consist of sets of ten-position manually controlled switches with each switch representing a decimal time digit. Since simulated time is divided into millionths of an hour, these switches are preset to the millionth interval, say 003162, in which an event is to occur. Information from each switch is transmitted to the *time comparator*, Fig. 12. Also transmitted to the time comparator, from the master time counter, is the simulated machine time, say at present, 003159. When the master time counter advances to a time 003162 which coincides with the detector time, the time comparator generates a stop-time signal to stop the clock. An action signal, associated with this particular time counter in coincidence is also lighted. The operator, after taking the appropriate action, resets the time counter to a future time—the time of the next waiting item in the category—and depresses the start time key. Checking circuits prevent advancement of the clock should the time detector accidentally be set to a time value which has already passed in the throwdown run, in the example, to a value less than 003162.

Three time detectors are provided in the throwdown machine. One each is used for setting originating times of subscriber starts and of incoming calls. The third is used for releasing held items. Since the holding times of these items are measured in hundreds of time units the last two digits of the time interval are dropped, and only four switches are required.

It has been mentioned that the ratio of basic machine (clock) time to real time is in the order of 70 to 1. However, in operation, the flow of time is halted frequently to permit actions by the operators. The average interval between stops in the traffic runs which have been processed is less than 10 time units. Thus the machine time is only a small fraction of the time consumed in processing a traffic sample. The ratio of total processing time to real time has turned out in practice to be between 1000 and 2000 to 1 depending on the nature of the traffic sample being tested.

GENERAL PLAN OF THE CONTROL CIRCUITS

The major part of the throwdown machine circuitry is associated with the marker sequence and timing controls and the gate preferencing arrangement. The circuit plan followed in these two cases will be briefly described in order to illustrate how the throwdown functions were implemented.

The gate circuits simulate the action of the marker connector circuits of the No. 5 crossbar system which control the access of line link frames and registers to the markers. These circuits assign traffic to idle markers according to the preference rules used in the actual system. The circuits resemble corresponding circuits of the system. They employ two relays per connector and one relay per marker and are arranged with cross-connection terminals so that the preference order and number of connectors can be varied as required.

Each call handled by a marker consists of a series of events occurring in time sequence with time intervals between events corresponding to the "work time" consumed by the marker in performing required functions. The sequence of events is not fixed at the start of a particular call but may be altered from stage to stage depending on the particular busy and idle conditions encountered. A block diagram of the control circuits used for simulating this action is shown in Fig. 12. They consist of a number of individual circuits provided on the basis of one per marker together with common circuits whose use is shared by all markers.

The fundamental plan is based on the use of two rotary stepping

switches. One of these, the sequence switch, determines the sequence in which events take place, while the other, the timer switch, measures work time between events. In the throwdown machine each possible event which may occur in setting up a call is represented by a position on the sequence switch. Circuits through the switch cause action signals for the event represented by its position to be displayed to the operators. All of the events for one type of call are arranged in order on the switch terminals so that by omitting certain positions (events) all of the variations of this type of call can be represented. At the conclusion of an event the switch is directed to the position corresponding to the next appropriate event according to the setting of "memory" relays located in each marker unit. These relays operate at various stages of the call in response to key signals from the operators indicating the conditions they have encountered in attempting to respond to action signals. Thus, significant events are recorded in order to control the future progress of the call.

To provide for all types of calls it was found necessary to furnish three sequence switches for each marker. These switches have 22 positions and six arcs. Five arcs are used for displaying signals and for control while the sixth arc is used in conjunction with the timer switch to control the time at which signals are displayed. A timer switch is individual to each marker. It is a 22-position, six-arc switch arranged with auxiliary relays to count a maximum of 105 clock pulses. Terminals of the timer switch

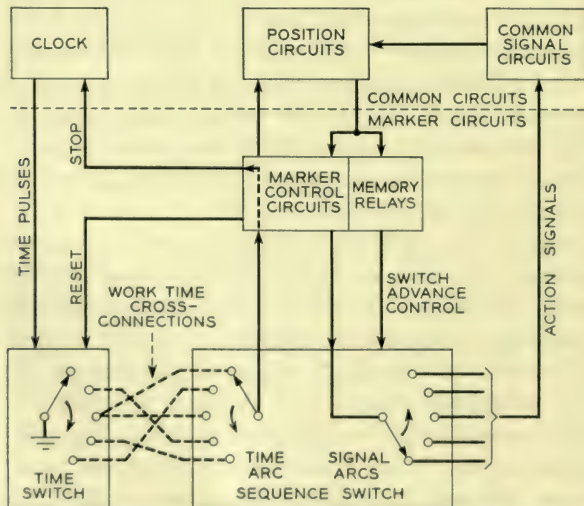


Fig. 12—Block diagram of marker progress circuits.

are cross-connected to the time arc of the sequence switch to fix the work time preceding the event represented by each sequence switch position.

In the general operating scheme, the sequence switch stands at a position representing the next event to take place. The timer switch is started from normal and counts clock pulses until it reaches the terminal cross-connected to the position on which the sequence switch stands. This initiates signals which stop the clock and cause action signals to be displayed. When the operators respond, the sequence switch advances to the terminal for the next event in the call, the time switch returns to normal and the clock restarts. The time switch then counts time units leading to the next event. Since several markers may be in use at the same time in different stages of their calls, two markers can reach an action point during the same clock pulse. A lockout circuit insures that only one marker at a time displays its action signal. At any time that a marker stops the clock, the timer switches of all other markers halt but continue their count when the clock is restarted. Thus, relative time relations are maintained while a true count of time consumed by an operating system is obtained.

The circuit action can be illustrated by a discussion of the events in a dial tone call. This call represents an attempt by a marker to establish a connection from an originating subscriber line to a dial register. The possible sequences of events are diagramed in Fig. 13. As indicated, this class of call employs eight sequence switch positions in addition to the normal position.

The call starts when the gate circuit assigns an idle marker to a call which has been originated at the proper time and placed in the gate. The assigned marker is prepared to process this type of call by the operation of an associated "class" key which selects and advances the sequence switch which carries the events of a dial tone call. Advance of the switch is from normal to Position 1 to control, at the proper time, signals for the first event, namely, seizure of a trunk link frame and selection of a register. The timer switch is set at zero and in a condition to step one terminal at a time in response to clock pulses. Terminal 1 of the sequence switch time arc is cross-connected to a terminal of the timer switch representing the marker work time between the time the marker is first seized and the time it attempts to seize a trunk link frame.

The operator in control of the marker now operates a START key and the clock starts pulsing, each pulse causing the marker timer switch to advance one step. When the specified work time has elapsed the timer switch reaches the terminal connected to Position 1 of the sequence switch. This passes a signal from the timer switch through the sequence

switch causing the clock to stop and the signals associated with Position 1 of the sequence switch to be displayed. Signals at the assignment position identify the particular marker and instruct the assignment operator to obtain a trunk link frame and a register. Three conditions may occur: (1) a trunk link frame and register may be idle and available, (2) a register may be idle but the frame on which it appears is busy, being held by another marker, or (3) all registers may be busy.

In condition (1) the assignment operator obtains the frame and register according to the operating procedures and operates an OK key at her position. This extinguishes her signals and passes a signal to the marker operator instructing her to record the present time in the space provided for trunk seizure on the call slip. The proper space is indicated by one of a row of lamps beside the call slip. The marker operator then operates the START key for this marker causing the marker timer to return to normal and the sequence switch to advance to Terminal 2. The clock starts if no other marker is waiting to display signals.

In condition (2) the assignment operator observes that there will be a delay in obtaining a trunk link frame. She inserts a connector cord for this marker in a "delay" jack and operates a DELAY key. This extinguishes her signals and passes a signal to the marker operator to record

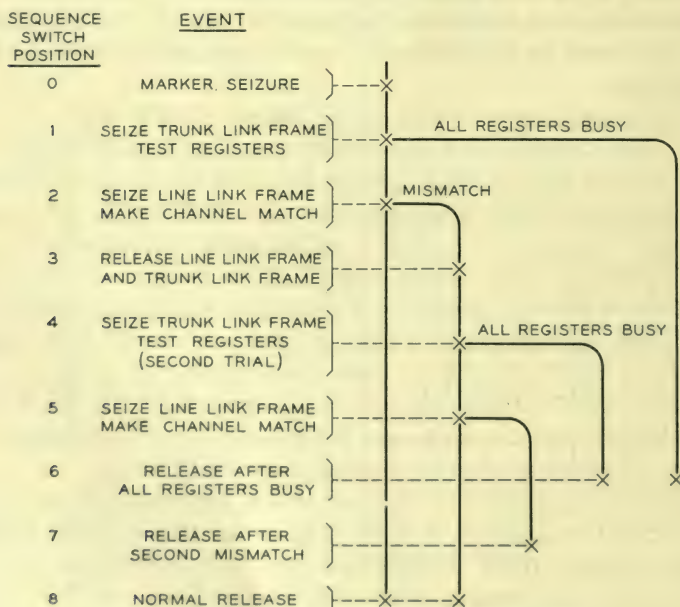


Fig. 13—Possible sequences of a dial tone call.

on the call slip that a delay has been encountered. The marker operator then operates the START key for this marker. This returns the timer to normal but since a delay condition has been established in the marker circuit by the actions of the assignment operator it does not advance the sequence switch but allows it to remain in Position 1. The marker in a delay condition does not permit its timer switch to step but removes the halting signal from the clock to allow time to advance and other markers to be served. All signals for this marker are extinguished during this time. As time advances some other marker will release a trunk link frame. This passes a signal through the delay jack to the delayed marker causing it to display again the signals requesting a trunk link frame and a register associated with sequence switch Position 1. The operators and circuits then proceed exactly as when these signals were first displayed.

In condition (3), the assignment operator observes that all registers are busy and operates her ALL BUSY key. This operates a memory relay in the marker circuit recording that the all busy condition has been encountered and that the future progress of the call should follow the sequence indicated in Fig. 13 by the side branch at Position 1. With all registers busy it is impossible for the marker to establish a connection. As the side branch shows, the alternative is to release and restore to normal. (Later attempts to serve this subscriber will be made until an idle register is obtained.) Thus, with the all busy condition recorded on the memory relays, the circuit will cause the sequence switch to advance, running over positions representing intermediate actions and coming to rest in Position 6 where it is prepared to display signals for the release of the marker. The operation of the ALL BUSY key also started the clock and restored the timer switch to normal so that it could measure suitable work time before displaying release signals.

The call progresses through successive events in a manner similar to that described above. After obtaining a trunk link frame and a register as in condition (1), the sequence switch stands in Position 2 while the timer switch counts work time preceding a request for a line link frame. During this time other markers may request service causing the clock to halt, interrupting the advance of time in all circuits until the operators have completed the required actions. After this marker has counted the specified time, a signal is passed through Position 2 of the sequence switch causing a request for a line link frame to be displayed at the marker position. A delay is handled as before, the marker waiting until the busy frame is released by some other marker. When the frame is obtained, a signal at the match position requests that operator to match

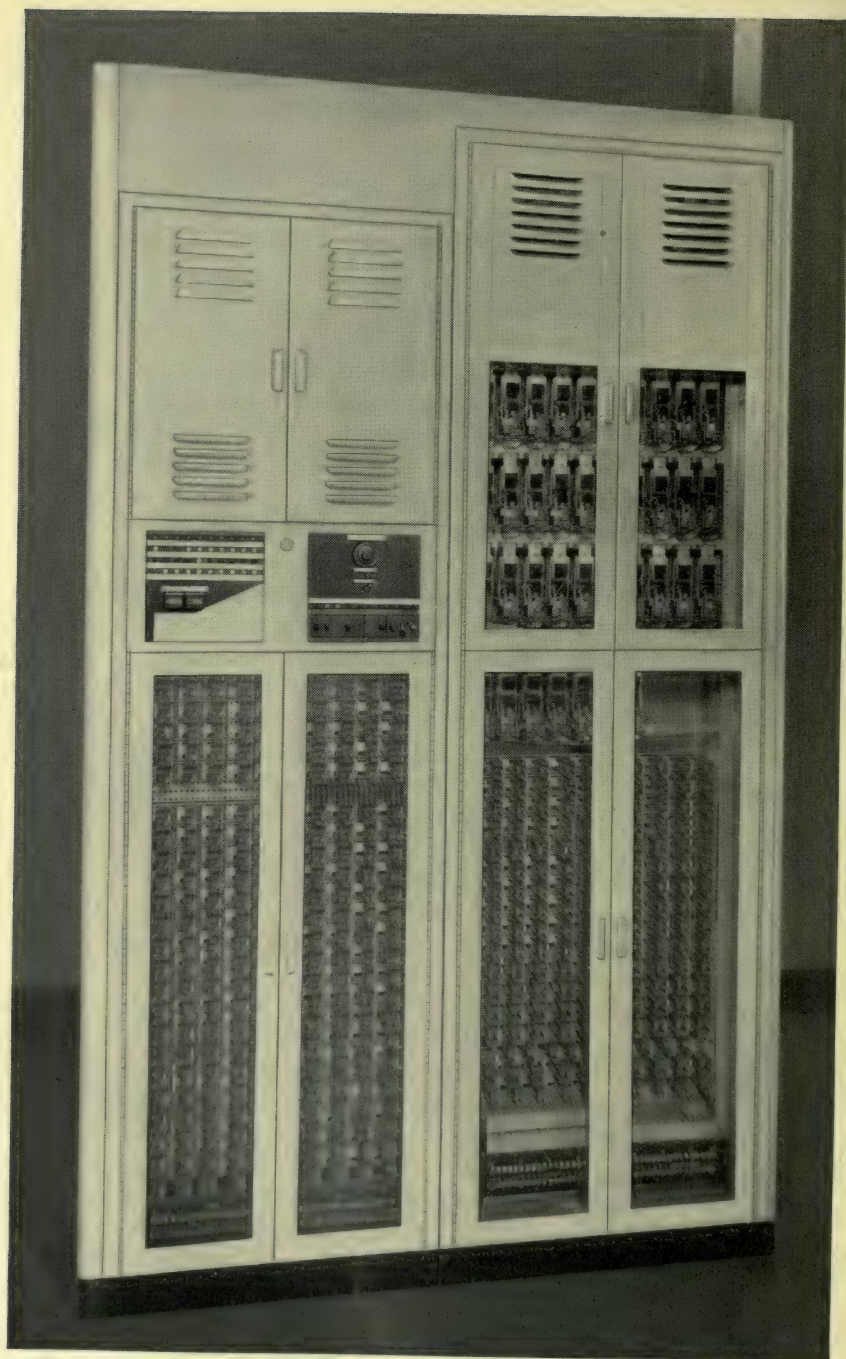


Fig. 14—Relay and switch cabinets.

a channel before allowing time to advance. In case of a mismatch the match operator depresses a MISMATCH key. This operates a memryo relay which will cause the marker to follow the alternate sequence indicated by the side branch at POSITION 2 in Fig. 13. As a safety precaution, the MISMATCH key is made effective only at the time a request for a match is made so that accidental operation at other times will not disturb the circuit action. With a mismatch recorded, the sequence switch advances to Position 3 and at the proper time displays signals to release the line link frame, trunk link frame and register. Thus the call advances with the possible alternates of all registers busy at Position 4 or a second mismatch at Position 5. The call proceeds to one of its possible conclusions where frames are released and the marker becomes idle.

The dial tone call which has been described is the least complex call handled by the machine. The intraoffice call which was diagramed in Fig. 4 is the most complex. It requires 22 sequence switch positions to represent events and may have 92 possible sequences depending on conditions encountered. To take care of variations in sequence in all calls a total of nine memory relays is provided in each marker.

The circuit equipment consists, largely, of telephone type electromagnetic relays and rotary stepping switches. Approximately 800 relays are used. The total number of switches is 57. Of these 47 are of the 22-position, 6-circuit type while the remainder are of the 44-position, 3-circuit type. This equipment is mounted in the two cabinets shown in Fig. 14 and in additional units located within the operating positions. Time indications are given by four-digit message registers, approximately 40 being used. The random circuit consists of a gas tube counting ring with several control relays. Output indications are given by miniature neon lamps. Signals are given to the operators by telephone switchboard lamps, 822 being used. Manual equipment used by the operators in signaling to the circuits consists of keys, switches and telephone plugs and jacks. The machine contains 187 keys and switches, 60 cords equipped with plugs and 509 jacks.

PREPARATION FOR A THROWDOWN RUN

There are two phases in the preparation for a throwdown run. One of these is the tentative engineering of an office of the size to be tested. The other is the preparation of data to represent traffic handled by this office. The first of these follows rather closely the general procedure that would be used in planning the installation of a new switching office.

The chief difference is that preliminary decisions concerning the size and general characteristics of the office to be tested will be made. For example it may be decided to test an office in the twenty line link frame size range serving mixed business and residential subscribers with a high percentage of calls completed within the office. Based on general knowledge of subscriber behavior, the number of subscribers necessary to present a suitable traffic load to this number of link frames will be determined. Values for average holding time and calling rate associated with these subscribers must also be developed. These may, in part, be determined from estimates or specific knowledge of the traffic capacity of the line link frames, taking care that the figures are typical of such a group of subscribers as determined by field observations. Since we are usually concerned with determining the maximum safe capacity of the system, full load or overload conditions will be assumed and the usual margins for future growth considered in engineering an actual office will be omitted. Decisions will also be made as to the number of other offices to which this office has trunks and the percentages of the total traffic originated and terminated in each of these offices.

Having made these preliminary assumptions concerning the nature and environment of the office to be tested, the office is then engineered according to the best available information. The numbers of registers and markers are determined and arranged in connector groups according to standard procedures. The sizes of the trunk groups to various connecting offices are determined and the placement of trunks on the trunk link frames chosen according to the usual practice. All similar factors concerning quantities and arrangement of equipment are determined. The throwdown machine is then set up to simulate this office. This will involve crossconnections in the gate circuits and arrangement and designation of the facilities provided for keeping the busy-idle records of such items as lines, trunks, registers and links.

The second phase in the preparation is to produce data representing calls presented to the system during the time interval to be studied. This is accomplished by choosing a random number for each call. This number must contain a sufficient number of digits to specify all the pertinent data necessary to describe the call. These digits are assigned to represent certain factors. For example, the first six digits represent the time of origination. The next two digits specify the type of call. This is done on a percentage basis. For example, if 40 per cent of the traffic is to be locally completed, the numbers 00 through 39 in these places would indicate an intraoffice call, if 25 per cent is to be outgoing to other offices the numbers 40 through 64 would indicate this type of

call, and so on for the remaining types of calls. If divisions of less than one per cent are to be made, three or more digits could be used for this purpose.

The meanings of certain of the remaining digits will depend upon the type of call. If the call is originated within the office, the next five digits will give the identity of the calling line in terms of its frame location. If the call is incoming from another office certain of these five digits will be used to specify the office of origination and the trunk number. This again is on a percentage basis depending on the fraction of total incoming traffic expected from each office. Five other digits give the identity of the called line if the call is completed within the office; if outgoing they specify the terminating office. Other digits give the percentage of calls which will result in partial dialing by the calling subscriber. Since, as will be discussed later, there is a possibility that these will be re-originated later as good calls, all of the information for a good call is also determined for these calls. Additional random numbers determine various other aspects of the call such as the identity of the number group which will be used. As a suggestion to those who would undertake a throwdown study, it is advisable to include a number of surplus "utility" digits in the original random number. It invariably happens that some factor is overlooked in the initial planning or arises during the course of a test and these digits can be used in making decisions in these cases.

It should be noted that the random circuit is used for making certain random decisions in the course of a call at the time that a need for these decisions arises. For example, the holding time of an established connection and the probability of the called subscriber not answering is determined by this circuit. This circuit could have been eliminated by including digits in the original number for every possible situation of this type which might be encountered. This would cause much useless work in preparing the data since all situations are not encountered on every call.

A quantity of random numbers must be drawn to provide the desired load on the office. This is not a simple process of determining the number of calls expected during a given period and drawing this number of random numbers. One factor is that in generating data by the above procedures it will be found that certain numbers will represent calls originated by lines which are busy at the indicated time on a connection established previously. The number of such cases can be estimated from the expected number of busy lines in the office and a corresponding number of additional numbers drawn. When this situation is encountered

in the course of the run the call is discarded. The other important factor is a "feedback" effect due to the calls meeting a situation which prevents successful completion and the probability that these calls are originated later.

A simple example is where the called line is found busy. As previously mentioned, there is then a possibility that subsequent attempts will be made at a later time. All attempts make use of circuits and control equipment and should be considered in determining the load capacity of the office. Other situations which produce this effect are illustrated in Fig. 15. Lines entering the figure at the left represent classes of calls which enter the system. A certain number of calls will be partial dials, no dials and false starts regardless of the performance of the system at the time the call is originated. The partial dials represent cases where the subscriber makes an error in dialing or does not wait for dial tone and dials a digit before he is connected to a register. These may be abandoned before the register obtains a marker or may "time out" in the register and be connected to an overflow tone trunk. In either case the subscriber may re-originate the call later. The probability of re-origination has been estimated from field data and the dotted lines in the figure represent these reoriginated calls.

In the throwdown machine the random circuit is used to determine which calls will re-originate and the elapsed time before the second at-

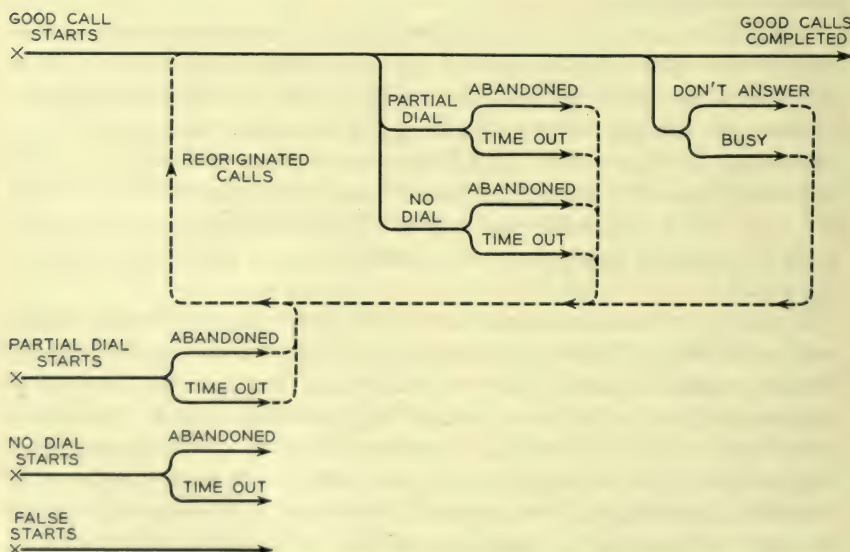


Fig. 15—Composition of the load on a crossbar office.

tempt. No dial starts represent such situations as a change of mind by the subscriber after lifting the receiver or accidentally removing the receiver and are not considered to be correlated with later trials. The same is true of false starts which are momentary start signals often hard to explain. A certain number of partial dial and no dial calls are the result of dial tone delays which may occur during heavy load conditions. These branch from the "good call" line in the figure and are due to the subscriber not waiting for dial tone and dialing part of the digits or even the entire number before being connected to a register. The probability of this occurrence will depend on the extent of dial tone delay. After dial tone delay on each call is known, successive uses of the random circuit determine whether the call is partial or no dial type and if the call is to be re-originated at a later time. Rough estimates of the quantity of this type of traffic could have been made and included in the original traffic data. However, since they depend on the performance of the system there is a tendency toward a "snowballing" effect and it was thought best to handle it as described in order to detect this effect.

It can be seen that the exact load on a system is difficult to estimate on the basis of initial starts. The procedure then is to make the best possible estimate of initial starts necessary to produce a given load, taking into consideration all important known factors and, at the conclusion of a run, make a count to determine the exact number of calls of various types handled.

When the data for the various calls have been determined from the random numbers, the pertinent information must be transcribed on the call slips. For most calls originated in the office this will require two call slips. One of these is for the dial tone stage of the call and is used in establishing a connection to a register. The other represents the later stage of the call where the register connects to a marker after dialing is completed and an attempt is made to establish an intraoffice or outgoing connection. The initial time of origination and the calling line number will be recorded on the dial tone slip. The slip for the second stage of the call will carry the calling line number and the called line or outgoing trunk number, but will not carry a time of origination. This is a function of dialing time and is determined at the conclusion of the dial tone stage by the random circuit. It is recorded on the second call slip at that time. The two slips are associated by recording the serial number of the associated slip on the dial tone slip. Incoming call slips carry the origination time, the calling trunk, and the called line numbers. Call slips for partial dial and no dial initial starts carry information similar to that on the dial tone slips.

In preparation for a run all slips bearing initial time entries are stacked in order of their times of origination. Slips for the second stage of a call are kept in a separate stack. At the conclusion of a dial tone action when the time of origination of the second stage is determined, the associated call slip is located, the time is entered on this slip and it is inserted in proper order in the stack of originating calls. In the case of re-originated calls, the information from the old call slips is recopied on a new slip with the new time of origination and these slips are placed with waiting calls in the proper time order.

DETAILED DESCRIPTION OF EQUIPMENT AND FUNCTIONS OF THE OPERATING POSITIONS

THE ORIGINATING POSITION

The originating position, in relationship to the rest of the throwdown machine, appears to the extreme right of Figs. 1 and 10. Detailed views of the two wings of the position are shown on Figs. 16 and 17. The principal features of this position are arrays of jacks and wooden pegs representing subscriber lines and incoming trunks, together with associated time counters and detectors. The chief function of the originating operator is to enter all calls into the system at appropriate times. This requires that the stacks of call slips, visible in Figs. 16 and 17, be held at this position. These call slips are arranged in order of their originating times, with latest time on top, and carry the originating line identification number so that line peg and call slip can be associated when the call is to start.

The line array, split into two wings, contains jacks and pegs for all subscriber lines in the office. The jacks which are simply holes with no electrical function, are arranged in a coordinate grid to assist in quick location. Twenty frames are provided, each holding 500 lines for a total of 10,000 lines.

The array is divided into 20 horizontal sections, running across the two wings, each representing a line link frame. Each frame is divided vertically into 10 subgroups of 50 lines, each representing a horizontal group. Since the directory number assigned to a subscriber line in a crossbar office is purely arbitrary and has no physical significance, it is not used in the present case for line identification. Rather, an equipment number, which represents the location of the line on a line link frame, is used. This is a five digit number, stamped on the peg, and made up

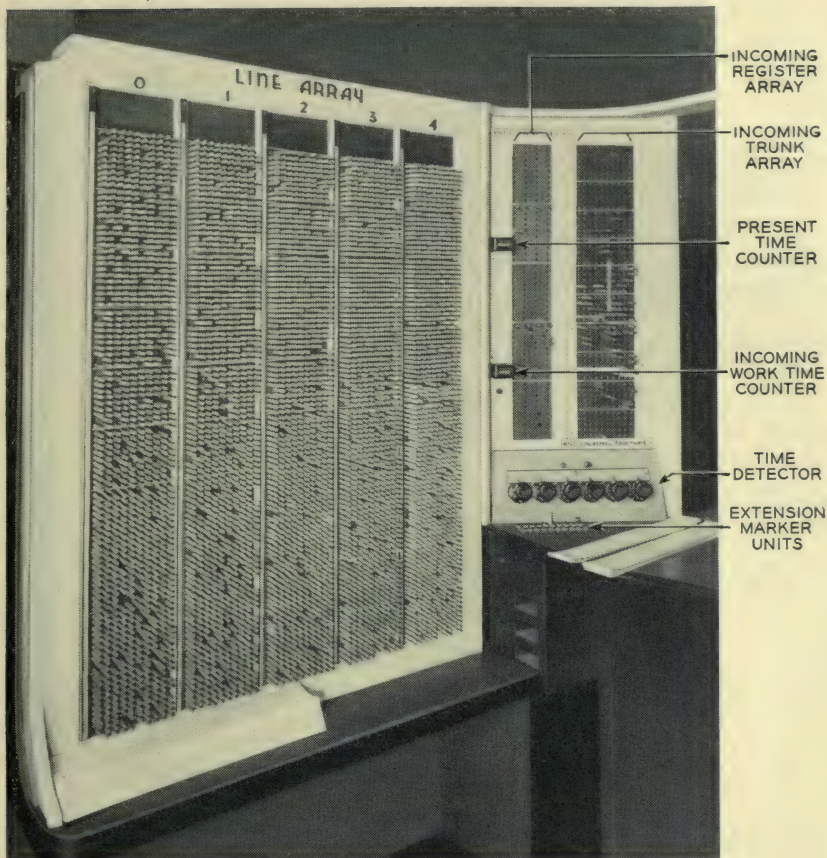


Fig. 16—Originating position—left side.

as follows:

HG—LLF—L

where HG—horizontal group No. 0-9

LLF—line link frame No. 00-19

L—line No. 00-49

The right upper five lines in each horizontal group of a frame (lines 45-49) are reserved for PBX use. A PBX can be assigned to line jacks occupying the same relative location in a vertical row (20 line link frames—same horizontal group). The pegs in a PBX group are marked with a distinctive color to facilitate hunting.

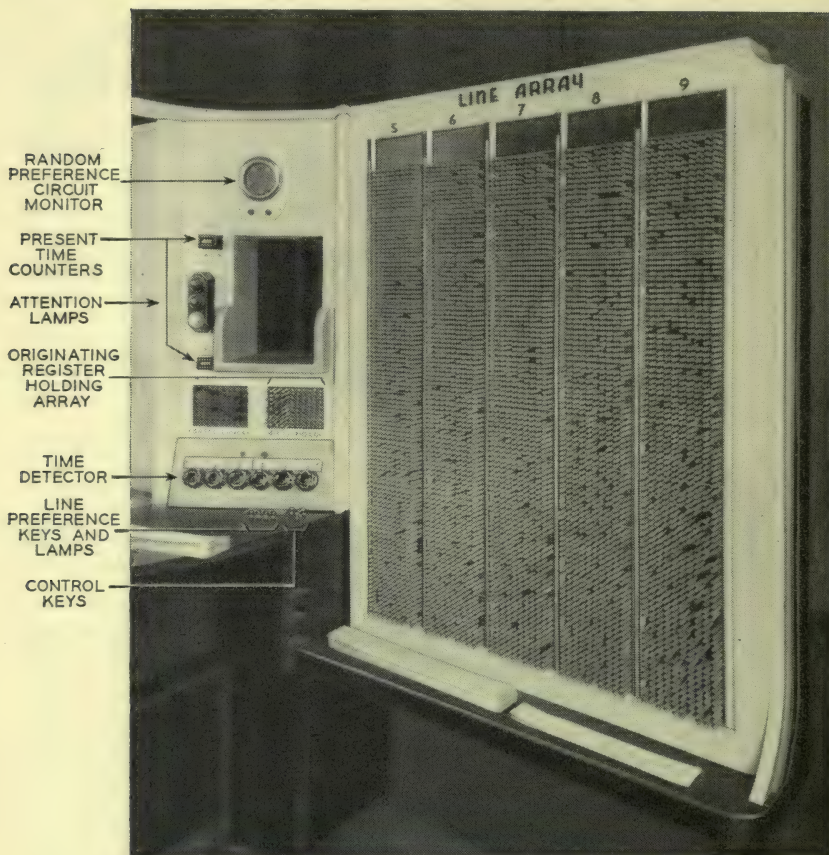


Fig. 17—Originating position—right side.

Line pegs are removed from the array while they are occupied on a call. The jack thus remains empty during the busy interval. The originating operator utilizes a line peg and a call slip in originating each dial tone call.

The incoming trunk and register arrays are located on the left column at this position as shown on Fig. 16. The incoming trunk array consists of pegs and jacks arranged in accordance with the trunk groups. The primary horizontal grouping is assigned a route number which is used only for identification purposes. The horizontal rows within a route represent the number of the trunk link frame switch on which the trunk is located. Vertical rows represent trunk link frames. Provision is made for a maximum of 400 incoming trunks. Trunk identification is a four-

digit number made up as follows:

R—TLF—SW

where R—route No. 10, 11, 12, 13

TLF—trunk link frame No. 0-9

SW—switch No. 0-9

Jacks representing these same incoming trunks also appear at the assignment position.

The incoming registers consist of short sleeves which can fit over the trunk peg. When a call is originated by a particular trunk as indicated by the time on a waiting incoming call slip, the trunk is associated with the correct type of incoming register by slipping the sleeve corresponding to the latter over the trunk peg. The peg and sleeve then accompany the call slip into the system. Appropriate times for each action are entered on the call slip.

The incoming register sleeves are held on arrays adjacent to the trunks. They are identified by a three-digit number:

CONN—REG

where CONN—marker connector No. 0 or 1

REG—register No. 00-19

Waiting holes are provided at the register array for holding trunks if all registers are busy.

MARKER CONNECTOR OR GATE

This is a bridging position between the originating position and marker position. Through it must pass all call slips and pegs requiring marker service. The originating operator places call slips and pegs in the gate from one side and the marker operator removes them at the other.

The gate, shown on Fig. 18, provides jacks for pegs and slots for call slips arranged in blocks corresponding to the individual marker connectors for line link frames, originating registers and incoming registers. It is divided into two sections, a storage section in which calls wait for an idle connector, and an active section in which calls wait for an idle marker and removal by the marker operator. An associated relay circuit controls the flow of traffic through the gate, simulating actual No. 5 operation, and indicates by lamp the appropriate action.

When a call slip is ready to enter the gate, the originating operator obtains the corresponding line or register peg and places the peg and slip in a jack and slot, respectively, in the correct connector block of the

storage section of the gate. The connector block number is ascertained from the peg identification or from the call slip. If the connector is free a CONN READY (R) lamp lights to indicate that the peg and slip should be advanced to the connector block in the active section of the gate. If the connector is not ready, the lamp will light at some subsequent time and the connector delay time will be noted on the call slip. If several calls pile up in a particular storage connector block, means are provided for maintaining correct preference for entry into the active section in the case of registers. With line originations, a random line preference circuit under key control at the right side of the originating position key shelf picks, on a random basis, the next entry into the active section of the gate (except in the case of preferred lines whose pegs will have a distinctive marking) and which have precedence over ordinary lines. Keys for a choice of one out of two, three, four or five are provided. With

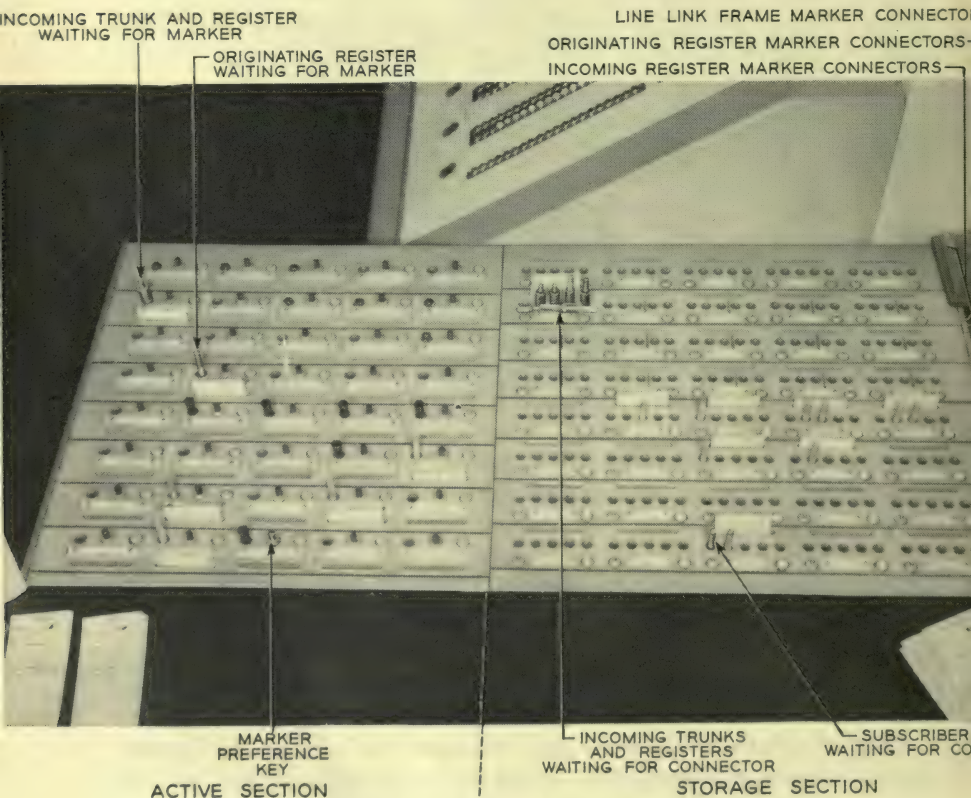


Fig. 18—The gate for control of marker connector preference.

register originations, the operator controls the preference with the assistance of a lamp signal per connector and a written register preference record.

Once in the active section of the gate, a call awaits an idle marker. When correct preference conditions are met, a CONN CLOSED (c) lamp lights, together with an attention lamp, to indicate to the marker operator which is the next call to be served. The marker operator depresses the MKR PREF key in the connector block to lock in a signal at the marker position indicating the marker to be used on this call. The operator then removes the call slip and peg from the gate and inserts them in the correct unit of the marker position. The match ticket portion of the call slip is torn off and passed to the match operator. On originating register controlled calls, the register match ticket will have been attached to the call slip and this also is passed to the match operator.

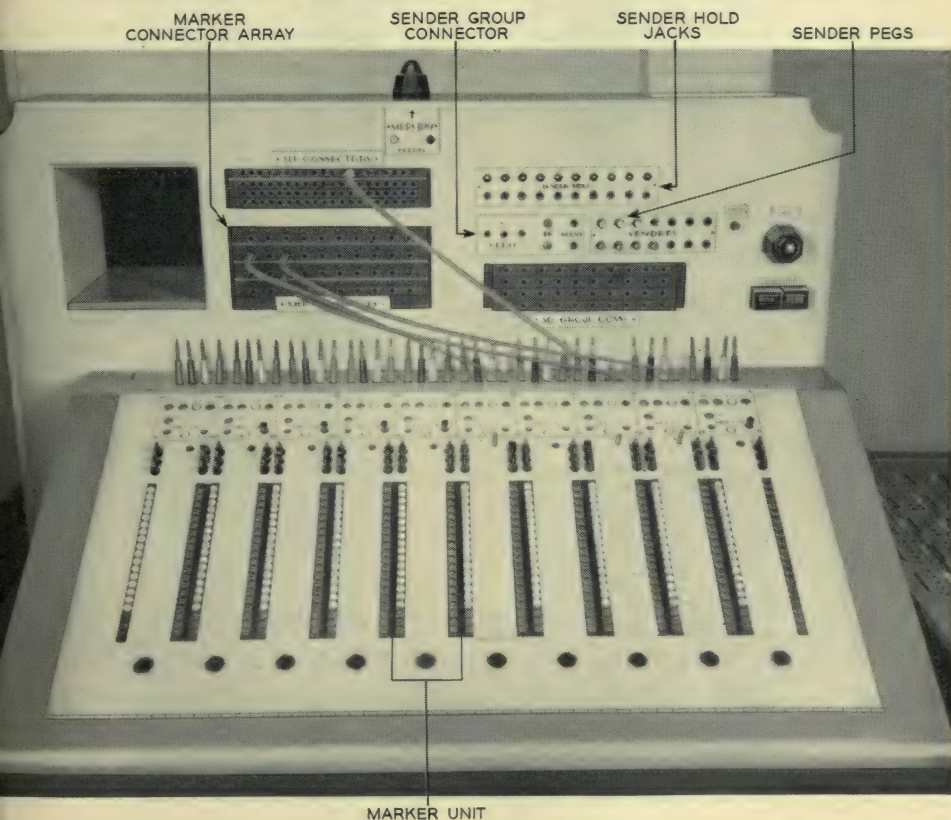


Fig. 19—Marker position.

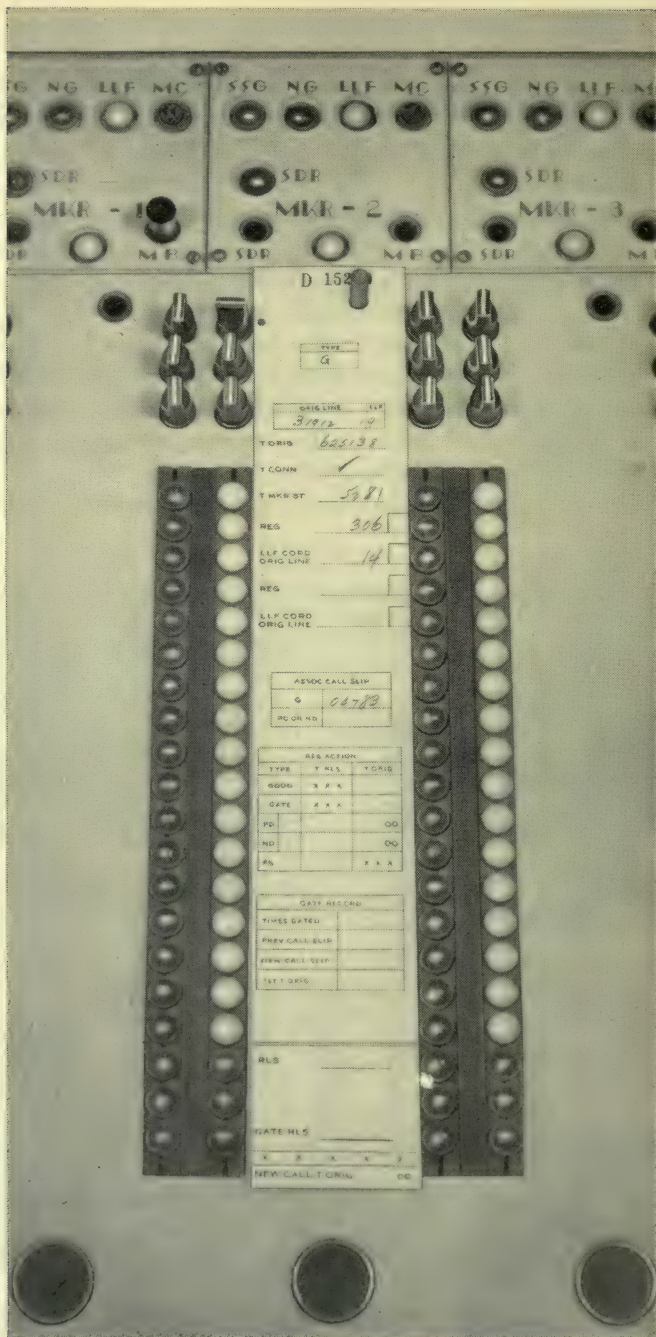


Fig. 20—Call slip in marker unit.

The circuit associated with the gate assigns all calls and markers according to No. 5 preference arrangements and provides the correct gating action. This preference is set up on cross-connection field located within the position and may be changed to correspond to any desired system arrangement.

MARKER POSITION

A photograph of the marker position is shown on Fig. 19. The individual marker units, of which ten are provided, are disposed on the sloping work panel. On the array panel are jack arrays representing the line link frames, the number groups, outgoing sender subgroups and senders, and the marker connectors. Connections to these jacks are made by means of plugs and cords located at the top of each marker unit.

The call slip for dial tone class of call is shown inserted in a marker unit on Fig. 20. At the top of the slip, in the TYPE and ORIG LINE boxes and opposite T ORIG and T CONN, are typical entries as made by the originating operator. The marker operator has also made three time entries.

The heavy dot at the top edge of the call slip indicates which of the six class keys to turn. The class keys condition the marker circuit to handle each type of call correctly. The hole at the top of the slip lines up with a jack and is used to store temporarily the line peg (register or incoming trunk peg in the case of other types of call) as received from the gate; it also serves to locate and hold the call slip in the correct position.

On the lower portion of the slip are spaces corresponding to all marker actions requiring association with other frames or circuits. These spaces line up with lamps on the marker unit which signal when and what action should be taken and indicate whether or not the time should be entered on the call slip. If a delay is encountered at any point, a check is put in the associated delay block to assist in subsequent computations involving the call slip. The left hand row of lamps lights to indicate when a direct action should be taken and a time written down. For example, when the third lamp from the left hand top lights (opposite LLF CORD) the operator places the line link frame cord at the top of the unit in the correct jack in the LLF connector array, and writes down current time in the space opposite the lamp. If the LLF jack is occupied by another marker, the operator plugs the LLF cord in a preference delay jack associated with the line link frame jack and places a check mark in the delay block.

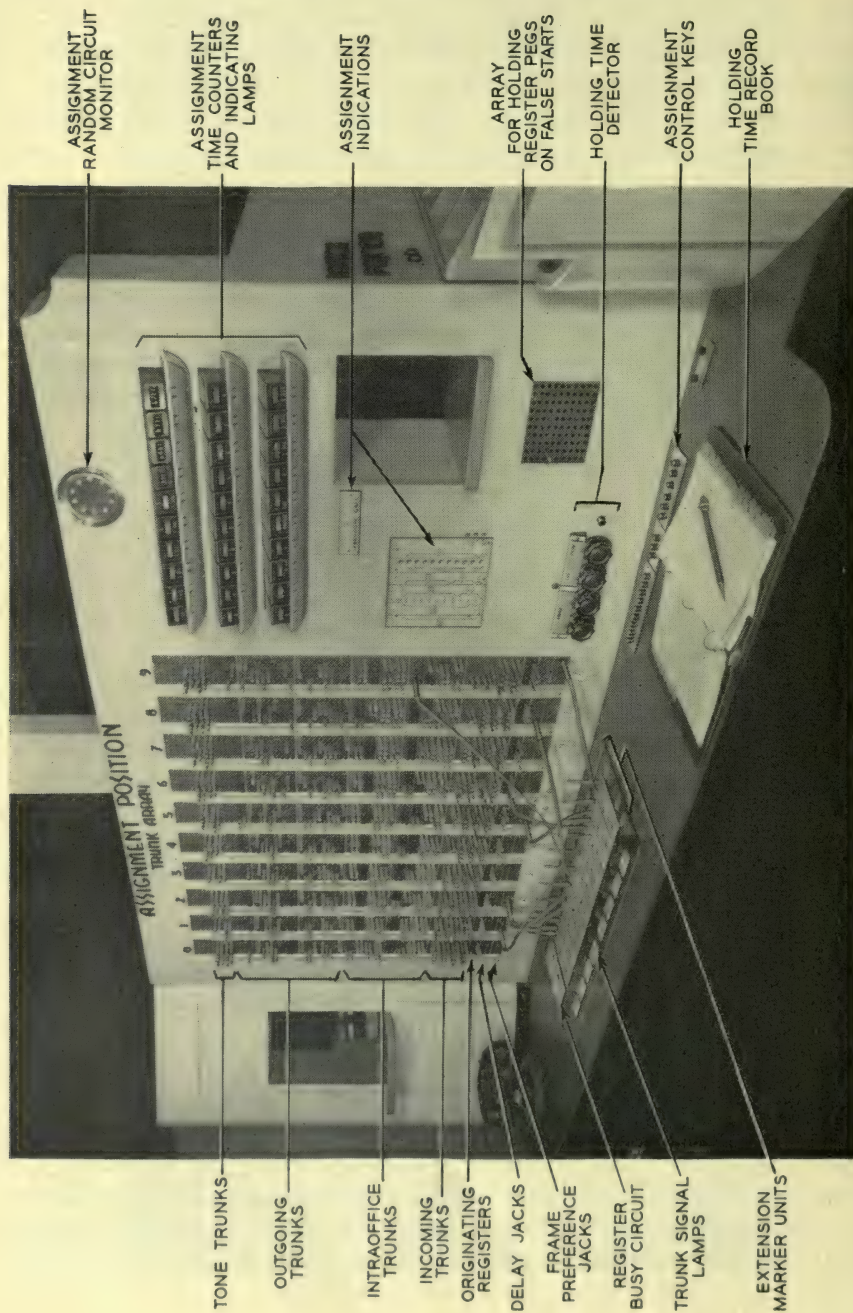


Fig. 21—Assignment position.

As discussed previously, certain marker functions are located at the assignment, originating and match positions. The choice of trunk link frames and trunks, for example, is made at the assignment position. Therefore certain lamps at the marker position light only to signal the writing down of a time. This is true of the second lamp from top left which indicates register seizure. The lamps in the right-hand row are also used in conjunction with actions at other positions. They light only to indicate that a delay should be checked in the corresponding box.

At the time of marker release, both left and right lamps light opposite one of the release categories. The specific type of release determines partially the further disposition of the call.

The START key at the bottom of each marker unit is pressed after completing each action called for by the sequence lamps. Operation of this key puts out the lamp and permits the circuit to advance.

ASSIGNMENT POSITION

The assignment position, shown on Fig. 21, includes the trunk array, the register false start holding array, the traffic assignment equipment, the holding time book and counters, and individuals units representing extensions of the markers. The chief function of the assignment operator is to test and choose trunks on each call, to determine the disposition of certain calls which encounter excessive dial tone or busy conditions, and to ascertain holding times of originating registers and trunks.

The trunk array consists of jacks (holes) representing all the trunks, and jacks and pegs representing the registers. Line pegs are held in the jacks during conversation time to mark the trunks busy. The array is divided into vertical sections representing trunk link frames. The trunk groups are disposed in horizontal rows so that trunk jacks of each group or route occupy the same relative position in each frame section. Thus a trunk-hunting action consists of picking a horizontal level in the array and searching along the level to the first idle trunk as indicated by an empty jack. The array provides more jacks than the total number of trunks so that most small trunk groups can be set up on individual horizontal levels.

The trunk identification is a three or four digit number composed as follows:

R—TLF—SW

where R is route number 0-99

TLF is trunk link frame 0-9

SW is switch number 0-9

The route number in some cases is required only for location purposes since many trunk groups are identified by name. It is assigned to a common group of trunks, maximum 100, or 10 per frame. However, in the case of outgoing trunks, the route number is required to identify the specific trunk group desired. The route number for originating registers is the same as the marker connector number for each register and is unusual in that any route number from 0-9 represents the same register group.

The register jacks occupy the lower trunk level on the array and are supplied with pegs which are used to originate the second half of outgoing and intraoffice calls. During dialing time, the associated line peg occupies the register jack to mark it busy. The next higher level on the array is the group of incoming trunks. These are furnished on a two-jack per trunk basis and are, in effect, multiples of the incoming trunk array at the originating position. The two jacks are required to hold the terminating line peg and the incoming trunk peg. The route number for these trunks is significant only for the function of location since it is not necessary for throwdown purposes to segregate incoming trunks into groups as it is with outgoing trunks.

The intraoffice trunks are divided into an A- and a B-group since more than 20 trunks per frame are required. The No. 5 marker can only test up to 20 trunks per frame at one time. The machine automatically allots calls between the two groups. Two jacks per trunk are required for originating and terminating line pegs. Unlike the incoming trunk jacks each jack of an intraoffice pair corresponds to a different switch location, although only one switch number is used for identification purposes.

The outgoing trunk groups are disposed above the intraoffice trunks. For the most part, these trunks will be in small groups, each with its own route number. When a call is set up to an outgoing trunk, it is necessary for the marker operator to inform the assignment operator of the correct route number. Several of these trunk groups can occupy the same horizontal level to conserve space. Only one jack per trunk is required to hold the originating line peg.

Tone trunks (busy, overflow, partial dial, no dial trunks) are at the top of the array. Only one jack per trunk is required.

Within the trunk link frame, the preference order in which a No. 5 marker tests trunks and the manner in which the order shifts from call to call is rather complex. In order to reduce the load on the assignment operator in trunk hunting, the actual trunk preference is approximated by reversing the direction of hunting within a frame group from call

to call on a substantially random basis. Thus, when the operator determines the trunk frame within which she will hunt, she observes the LEFT and RIGHT lamps below the array for the indication as to whether to hunt from left to right or vice versa.

TRAFFIC ASSIGNMENT

One of the important functions of the assignment operator is to determine the assignment of all calls at the time of marker release. For this purpose she is furnished with traffic assignment keys and indicators which appear to the right of her position. An adequate understanding of this feature requires a somewhat detailed explanation.

With the aid of the traffic assignment controls and indicators, shown to the right of Fig. 21, the assignment operator performs the following specific functions:

On Dial Tone Calls: The operator determines whether a call reaching a register should be classified as a good call (successful subscriber dialing), a PD call (partial dial—incomplete subscriber dialing) or an ND call (no dial—no subscriber dialing while connected to a register). A proportion of call slips are originally marked as PD or ND (and also FS or false start) and on these this determination need not be made. If the call is of the PD or ND type, the operator determines whether it should be subtypes PD1, PD2, PD3 or ND1, ND2, ND3. The subtype affects the assumed time until the subscriber abandons the call. If the call is classified as good type, the operator determines which of several dialing times should be used. If the call is classified as PD, ND or FS type, the operator determines when the call is abandoned and whether or not it is routed to a tone trunk.

On Calls Completed to a Subscriber: On calls completed via intraoffice, outgoing or incoming trunks, the operator determines whether the call is answered and which of ten holding times should be assigned for subscriber line and trunk.

On Calls Routed to a Tone Trunk: On calls which are routed to tone trunks or given a tone signal from the register, the operator determines the trunk or register holding time and whether and when the call is re-originated.

In making these determinations, the operator presses keys which cause a lamp to light either beneath a time counter or beside a designation strip. The time counter, set ahead of present time, indicates trunk release time, register return time, etc., while the designation strip classifies the calls. The determining factors include the magnitude of dial

tone delay, probability, and whether or not all registers are busy. Where dial tone delay is concerned, it is determined by comparing the originating time of a dial tone call slip with three dial tone delay time counters. These counters give present time minus 1, 2 and 9 seconds respectively so that matching the time of origination against them indicates whether the delay was < 1 sec., 1-2 sec., 2-9 sec. or > 9 sec.

The probability factor is obtained from a circuit which is capable of lighting one lamp out of ten, one out of three, etc. on a random basis when a key is depressed. By means of the circuit, calls can be assigned to various categories in correct proportion in accordance with the best available traffic information. Whether or not all registers are busy is indicated by the all register busy circuit controlled by the assignment operator.

If the assignment for a trunk or a register is that it be held for a period of time and then released, the assignment operator makes use of the holding time book as described later. If the assignment is for a register to return with a bid for a marker, or a new call to return into the system (after encountering busy, for example), the time is noted on the call slip and the latter is passed to the originating operator for subsequent action. For new call return time, a letter designation associated with the signal lamp is also entered on the call slip. The letter is carried forward on the new call slip. If subsequent attempts of this same line meet busy or overflow, the letter designation is used to identify the same category of return time instead of using the random circuit.

When a trunk call is set up, the trunk and one or two lines must be kept out of service for one of several fixed holding times. There are ten different assigned holding times with an equal likelihood of an established call falling into any one of them as determined by the traffic assignment circuit. False start, tone trunk and don't answer connections provide four additional holding times.

The holding time counters provided at the assignment position indicate present time plus a fixed holding time. Thus each counter gives the time at which a connection, set up at present time and assigned to that particular holding time, will release its elements back into service.

Holding time starts for a given call at the time the release lamp lights at the assignment position marker unit associated with that call. At such a time, the assignment operator obtains the one or two line pegs of the call and plugs them in the trunk jack or jacks (identified by the frame connector cord), noting at the same time the trunk number. The operator then depresses a key which causes the traffic assignment circuit to light a lamp under one of the holding time counters, thereby assign-

ing a release time. The trunk number and release time are recorded. At the end of the holding time, the operator removes the line peg or pegs from the trunk jacks and the pegs are returned to their home jacks. Presence of the line pegs in the trunk jacks during holding time marks the trunk busy.

The busy trunk numbers are recorded in a holding time book according to release times. Time units in the system represent millionths of an hour (0.0036 second). Each page of the holding time book covers 10,000 units of time and is divided into one hundred 100-unit blocks. An illustration of one page of the book is shown on Fig. 22.

Since the time unit is one millionth of an hour, each time unit in a one hour series can be represented by a six-digit number. Of these six digits, as far as the long holding times are concerned, the last two are unimportant since 100 units is only 0.36 second. If they are dropped, the first two digits of a four digit holding time release figure give the 10,000 time unit group or page number of the holding time book and the second two digits give the 100-block on the page. Trunk numbers are entered in the book on this basis with the page and block number obtained from the assigned holding time counter.

By this means the release times of all items appear in consecutive order in the holding time book. Holding time entries are always made several 100-blocks beyond the next release time, which eliminates con-



Fig. 22—Holding time book.

fusion. Following the simple numerical order of release times, the operator sets up the next release time on the holding time detector, similar to the time detectors at the originating position, which stops the system and signals when that time arrives. The operator returns the held items listed in that 100-block to normal, crosses out the block and sets up the next following time on the detector.

False start and register overflow release time entries are made in the same holding time book. A jack array is provided to hold the register peg until release time under these conditions.

Since the longest holding time is six minutes (100,000 units), the operator is never concerned with more than 11 pages at one times. With each page clearly labeled by a numbered margin tab, search time to make an entry is reduced to a minimum.

MATCH POSITION

The match operator performs the function of testing and making busy the switching channels through which each connection is set up. Her position includes ten marker units, each consisting of a group of lamps, and a set of files which hold the channel cards by means of which the channel records are kept. The position is shown on Fig. 23.

The No. 5 marker picks out a channel between a subscriber on a line link frame and a trunk on a trunk link frame by matching a line link, a junctor and a trunk link which are capable of being switched together to connect the two end points. A schematic of the system is shown on Fig. 24.

Each horizontal group of subscribers has direct access to a set of ten line links which connect to the ten junctor switches of the frame. Each link of the set can be given a number from 0 to 9 corresponding to the junctor switch on which it terminates.

Verticals on each junctor switch connect to junctors which are distributed over all the trunk link frames in the office. In a ten trunk link frame office, the ten verticals of each line link frame switch are distributed to all ten trunk link frames. This provides a set of ten junctors from each line link frame to each trunk link frame. If there are less than ten trunk link frames in the office, additional sets of junctors, perhaps comprising less than ten junctors per set, are provided between frames. The junctors connect between like-numbered switches and within a set bear the same number (0 to 9) as the switch.

The trunk links are similar to the line links except that twenty links connect from each trunk switch to the ten junctor switches. The twenty

links are subdivided into left and right sets of ten which connect to the left and right halves of the junctor switches respectively. Within each set, the links are numbered in accordance with the junctor switches on which they terminate. The junctor switches are split horizontally as shown on Fig. 24 to provide for twenty junctor connections per switch (one from each of twenty line link frames). Thus a set of junctors terminating on the left halves of a junctor switch must be matched with a left set of links.

It can be seen on Fig. 24 that when the three sets of links and junctors capable of connecting a trunk and a subscriber are matched, the two No. 0 links and the No. 0 junctor go together, the two No. 1 links and the No. 1 junctor go together, and so forth. The marker performs the

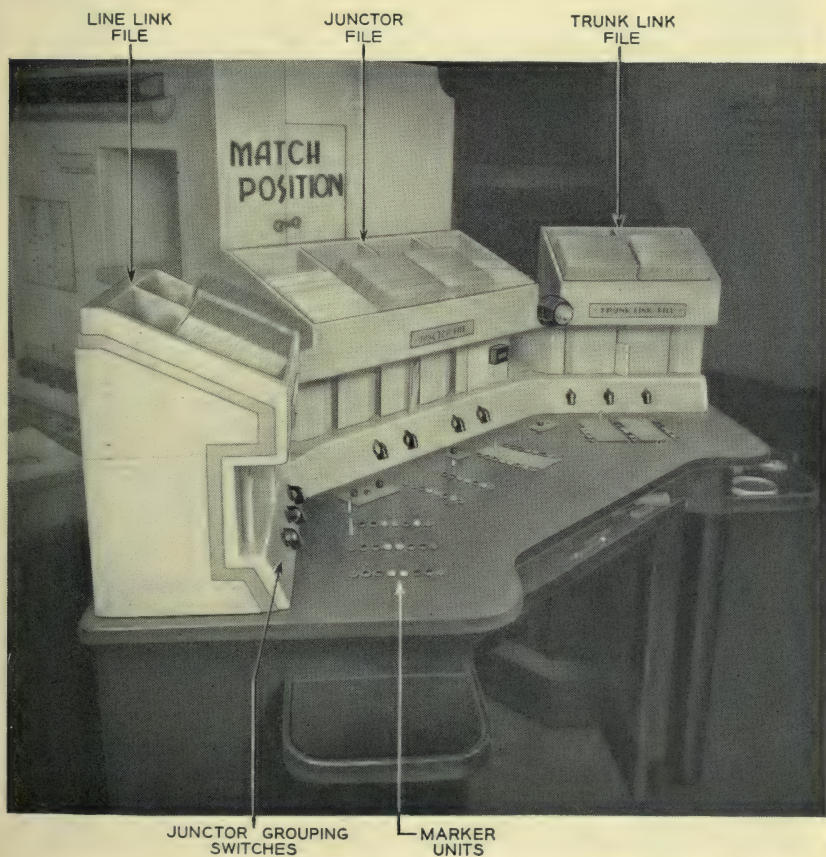


Fig. 23—Match position.

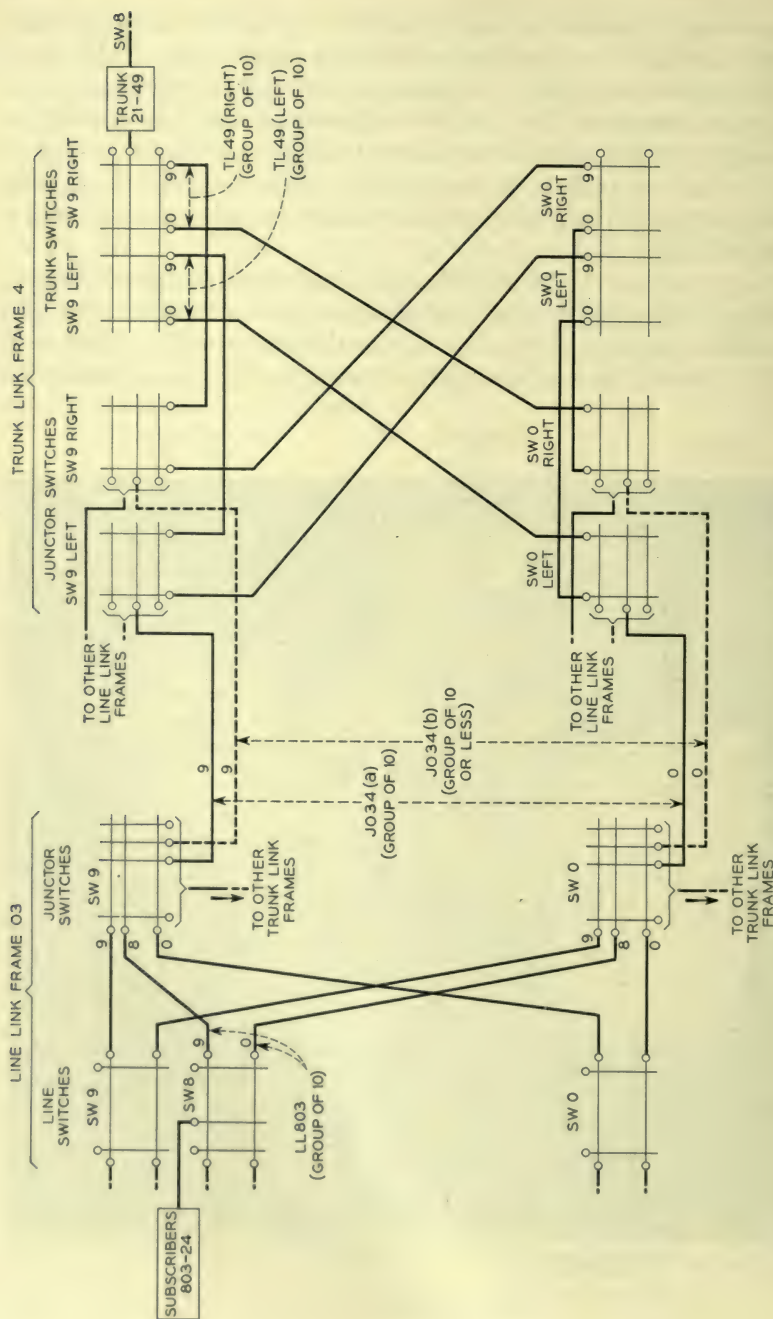


Fig. 24—Channel matching schematic.

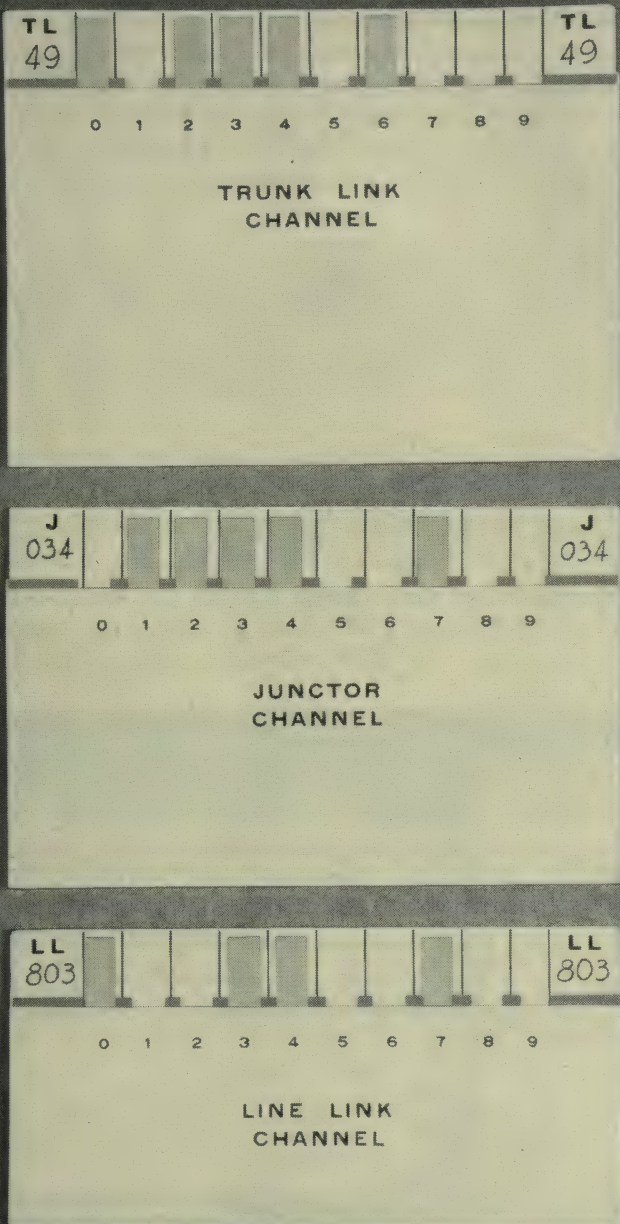


Fig. 25—Channel cards.

matching function by gaining access to a set of line links, a set of trunk links and a set of junctors on the basis of its knowledge of line and trunk location. It tests for three idle elements of like number within these sets and picks the lowest numbered idle channel available for making the connection. If it is impossible to match three idle elements and there is more than one set of junctors between frames, the marker will change the junctors and make a second attempt at matching. This marker function is performed by a six-stage allotting circuit which advances one step for every match operation. Depending upon the number of junctor subgroups, the allotter is arranged to rotate the first choice subgroup on each subsequent match operation and provide an alternate subgroup if the first match attempt fails. This system tends to equalize the use of junctors.

For use by the match operator, a channel card for each set of ten links or junctors is provided. Each card, as shown on Fig. 25, has ten pockets, one per link or junctor element, in which can be inserted a busy tab. When the three cards required for a particular connection are identified, they can be stacked as shown on Fig. 26 (note the difference in card size) and the idle channels are immediately obvious. In this case, channel 5 is the lowest available one and would be assigned to the call.

The identity of each card corresponding to a set of links or junctors

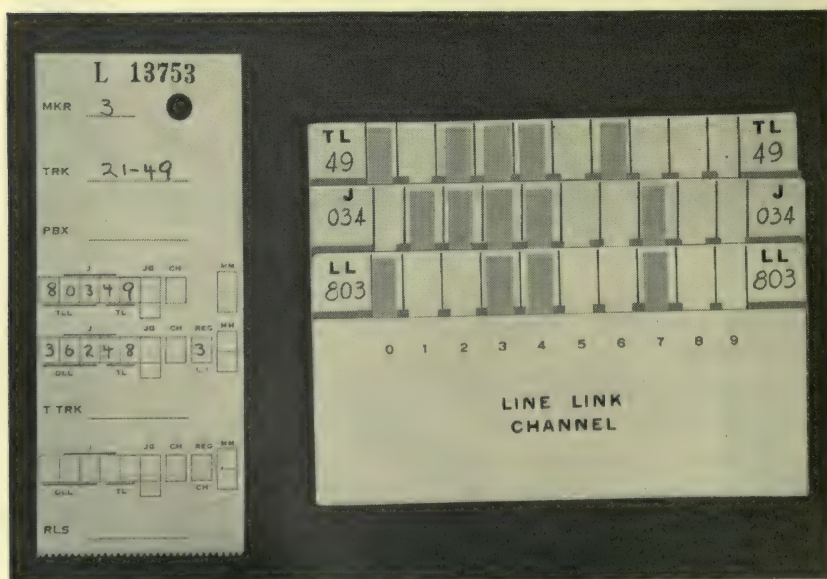


Fig. 26—Channel matching procedure.

is determined by switch and frame numbers. There are ten sets of line links per line link frame and a maximum of twenty frames. The identification of each set is a three digit number made up as follows:

SW—LLF

where SW—switch number (0-9)

LLF—frame number (00-19)

This provides for a total of 200 line link sets or cards.

The number of the line link set is incorporated in the line identification number so that from the latter can be determined immediately the particular line links available to the line. Note on Fig. 24 that line 803-24 must use the LL803 set of links (card shown on Figs. 25 and 26) for any connection.

There are twenty sets of trunk links per trunk link frame (ten left and ten right sets) and a maximum of ten frames. The identification is a two digit number plus a left or right indication. The number is composed as follows:

TLF—SW

where TLF—frame number (0-9)

SW—switch number (0-9)

and the left or right indication is, for convenience, one of two colors. This provides for a total of 200 trunk link sets or cards, 100 of each color. The link identification is included in each trunk number. Thus, on Fig. 24, trunks 21-49 must use trunk link set TL 49, either left or right.

The number and disposition of junctors are determined by the layout of line link and trunk link frames. In a 20 line link-10 trunk link frame office, there is one set of junctors between each pair of frames. For fewer frames there are more sets between frames to a maximum of five for a four-line link, two-trunk link frame office. The junctor sets or cards are identified by the two frame numbers involved, as

LLF—TLF

where LLF—Line Link Frame No. (00-19)

TLF—Trunk Link Frame No. (0-9)

If more than one set of junctors interconnect two frames, letters A to E are added to the base number. For a particular line and trunk, the junctor number is derived from a combination of the line and trunk numbers. For example, the essential parts of line number 803-24 and

trunk number 21-49 combine to give a number

80349

where the underlined digits represent the set of junctors (see Fig. 24).

In the general case, the junctors of a set may terminate on either the left or right half of a trunk link frame junctor switch. When the channel cards are made out preparatory to a throwdown study, the junctor numbers are assigned to cards of one of two colors (same as trunk link cards) depending upon whether it is left or right connection. Thus, in picking the three channel cards for a match, the choice of trunk link card depends upon, and must be the same as, the color of the junctor card.

Normally the channel cards are kept in filing sections at the match position as shown on Fig. 23. Before the matching operation is physically performed, the operator has available on a match ticket the line-trunk composite number. In the example shown on Fig. 26 this number is

80349

The digits 803 identify the line link card; the digits 034, the junctor card; and the digits 49, together with the junctor card color, identify the trunk link card. The operator removes the cards from the file, stacks them, and places them in a slot associated with the particular marker until the match signal is received. At that time, the operator picks the lowest numbered free channel in the stack, marks it busy and enters the channel number on the match ticket for record.

Each channel must be released at the same time that the line and trunk with which it is associated are removed from holding. Since release times for lines and trunks are entered on the assignment position holding time detector, this latter detector is used to signal release times to the match operator. The match operator maintains all her established match tickets in release time sequence with the earliest time on top of the pile. The lighting of a signal lamp indicates that the channel identified by the top ticket should be restored to normal.

A channel release condition which the match operator must recognize without a special signal occurs at the marker release time on intraoffice, outgoing, partial dial and no dial calls. At this time the register channel associated with the call must be dismissed. The operator will have received from the marker operator the register channel ticket involved and must restore the channel to normal before answering any new signal.

MASTER CONTROL PANEL

A master control panel for the throwdown machine is supplied on the relay cabinet shown on Fig. 14. This panel provides: means for turning the system on and off at the beginning and end of each day's operations; a centralized alarm indicating system; a bank of marker action lamps which indicate marker status during operation; and a present time counter with a units-to-seconds conversion scale.

The equipment operates on two battery supplies, one known as permanent battery and the other as day battery. Permanent battery is on continuously during a complete throwdown run in order to hold operated certain record relays. The day battery, however, is turned off during idle periods.

The equipment is arranged so that at the end of a working period all operator functions requested by signal lamps during the last working unit of time can be completed before the machine automatically stops. This is controlled by the DAY-NIGHT switch which is turned to the NIGHT position when it is desired to cease operation. The NIGHT lamp lights at this time. When the operators have extinguished the last signal lamp, the DAY POWER switch is turned to OFF.

Certain critical portions of the machine are provided with alarm circuits which automatically stop the machine and light lamps at the control panel. In most cases, the lighting of one or more of these lamps will require troubleshooting. When the trouble has been found, operation of the key associated with the lamp extinguishes the lamp and permits the machine to start again.

RESULTS OF THROWDOWN STUDIES

The throwdown machine has now been in operation for slightly less than four years. During this time 1383 seconds of equivalent central office time, divided among eleven runs, have been accumulated. The machine, of course, has not been in continuous operation, since the time required for preparation of a run and eventual analysis of output data exceeds the actual operating time for the run. In general, the same team of girls has handled both preparation and analysis of data and operation of the machine. Beyond this, there have been periods when no studies were in progress.

A detailed presentation of throwdown results is not properly within the scope of this article. However, a brief resume of the several runs with some mention of their primary objectives and typical results is necessary to conclude this picture of the throwdown machine.

It should be emphasized again that the principal function of a throw-down machine is to provide data, under controlled conditions, which can be used to develop and check a comprehensive theory of operation of a complex system. A secondary function is to study the reaction of the system to specific equipment or circuit arrangements. The No. 5 throwdown machine has been used for both purposes. All runs have furnished masses of statistical data which have been very useful in formulating traffic theories applicable to No. 5 crossbar. Some data, such as those on link matching, have a more general field of application.

All throwdown runs have employed a basic size office of twenty line link frames and ten trunk link frames, loaded by 9000 subscribers. On the first two runs, a traffic level designed to load these frames to their normal capacity was applied to the machine. This level can be called

TABLE III

Run	Duration	Quantity		System Loading	Primary Purpose of Run
		Markers	Orig. Registers		
	<i>Sec-ends</i>			<i>per cent</i>	
I	256	5	67	100	To load up machine and give normal load picture
II	216	5	59	100	To study effect of higher marker occupancy due to reduced number of registers
III	90	5	59	125	To study effect of 25 per cent overload
IVa	86	5	59	110	To establish response of system with original gate preference for comparison with run IVB
IVb	94	5	59	110	To provide data on response of system with reversed gate preference
V	65	6	65	—	This was an intermediate run in which traffic at high level was introduced in order to build up to system equilibrium at the 120 per cent load level
VI	108	6	65	120	To study system at high load level below saturation
VII	288	7	60	120	To study effect of situation where registers are more severe bottleneck than markers
VIIIabc	180	4	35	140	These three runs tested two proposed changes in the gate control circuit against the standard reversed preference arrangement. Identical traffic was used for each run

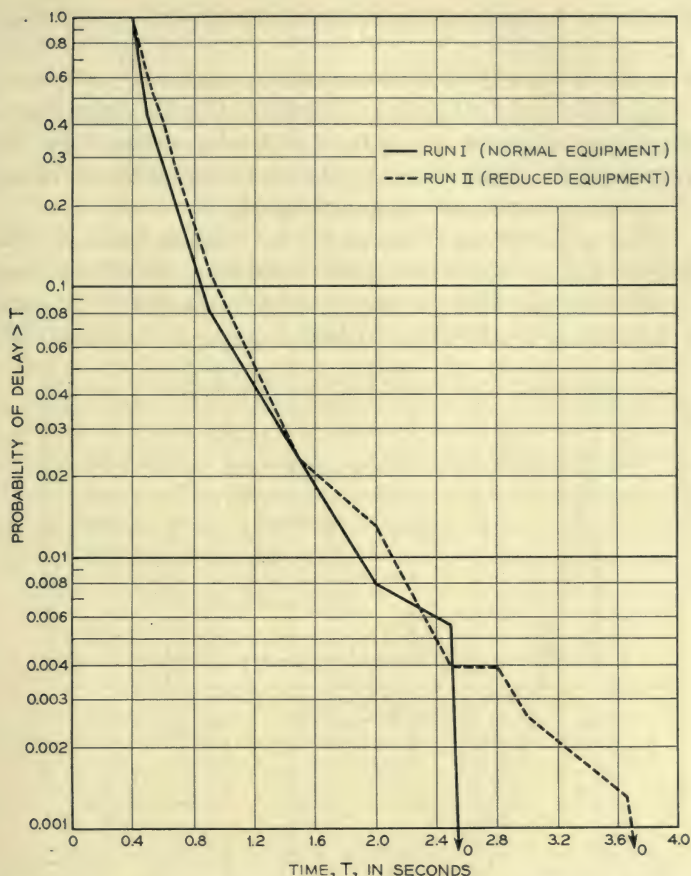


Fig. 27—Dial tone service with different equipment quantities as determined by the throwdown machine.

100 per cent load. In succeeding runs, an overload of varying amounts was utilized. In the several runs, the quantities of registers and markers were varied to obtain basic engineering data.

As the No. 5 system was originally engineered, the preference order in which markers were assigned by the gate control circuit to marker connectors during heavy loads was as follows: (1) line link frame marker connectors; (2) originating register marker connectors; (3) incoming register marker connectors. At a later date, this order was changed to put the line link frame marker connectors last in preference. In the discussion which follows, this arrangement will be known as "reversed preference."

The essential features of the throwdown runs to date are given in Table III.

The curves of Figs. 27 and 28 are representative of the type of data made available by the throwdown machine. Fig. 27 shows the overall dial tone service obtained during the first two runs. The slight degradation of service in Run II is caused by the reduction in number of registers. In both cases, however, service is very good.

Fig. 28 shows the spread of delays met by line link frames in obtaining a marker on dial tone calls during the same two runs. These curves are representative of the distributions of delays encountered at individual stages of handling a call. Other examples might be line link and trunk

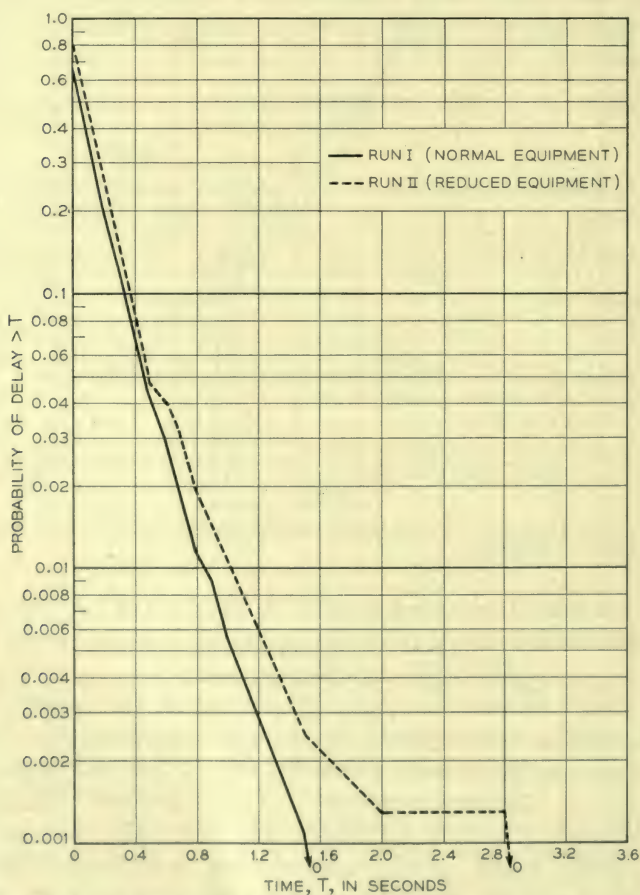


Fig. 28—Marker delays on dial tone calls with different equipment quantities as determined by the throwdown machine.

link frame delays. Taken together, these various delays determine overall grade of service.

An interesting example of the practical utility of the throwdown machine is furnished by the series of events which lead to the introduction of reversed gate preference in the No. 5 crossbar system. The changing quantities of line and register pegs in the gate position provide a graphic visual indication of the dynamic status of the system. After watching the gate for some time in the early overload runs, it was noticed that the register marker connectors were relatively less successful in gaining access to markers than the line link frame marker connectors. During all register busy periods, this reduced the call handling capacity of the system since registers were delayed in becoming available to waiting lines. The effect was compounded by wasting marker time in attempting to set up dial tone calls when no registers were idle.

It was felt that a change in gate preference, placing register marker connectors before line link frame marker connectors, would improve this. The new arrangement was tested and confirmed in throwdown run IV and is now a system standard.

Working Curves for Delayed Exponential Calls Served in Random Order

By ROGER I. WILKINSON

(Manuscript received December 19, 1952)

Working curves of delays for waiting calls served at random are given for a considerable range of loads and group sizes. Exponential holding time calls are assumed originating at random, and served by a simple group of paths. Results of a number of throwdown tests are given to illustrate the effect on call delays of several modes of service, and particularly of service on a random basis. For random service, these results verify the theory recently developed by J. Riordan; perhaps more interestingly they show the effects on delays of certain blends of queued and random service which approximate methods of handling delayed calls in practical use (such as gating and limited storage circuits). The use of random and queued delay theory is illustrated by a number of examples. To remind the reader that these results are not limited to telephony, department store and vehicular traffic problems are included.

A theory for predicting the delays which telephone calls (or other corresponding types of traffic such as vehicular, aircraft, people waiting in line, etc.) having exponentially distributed holding times would encounter when the delayed calls are served in a random order was published in a recent issue of this JOURNAL* by John Riordan. Mr Riordan's mathematical analysis involved a determination of the first several moments of the delay distributions. He then devised a method of combining elementary exponential curves in such a way as to satisfy the moments previously calculated.

Since a limited number of moments were used in the above determinations the curves derived are approximate only, but at the same time they are believed to be good approximations. The critical cases are those of paths carrying very heavy loads, in the occupancy ranges of $\alpha = 0.80$ or higher.

* Bell System Technical Journal, January, 1953, pages 100-119.

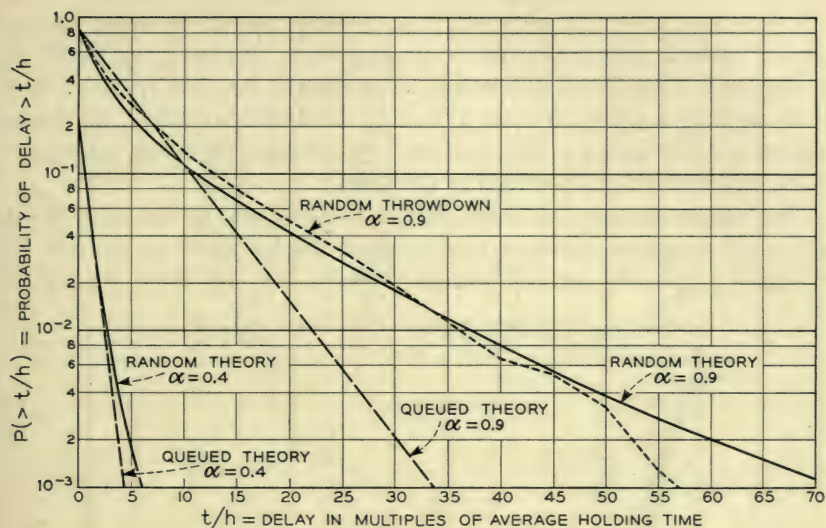


Fig. 1 — Distribution of delays. Theory versus throwdown, delayed calls handled at random, $c = 2$ paths, $\alpha = 0.90$, 3000 throwdown calls.

THROWDOWN CHECKS

Before calculating a field of curves for working purposes it was thought desirable to make at least a modest throwdown test, or traffic simulation, at these high occupancies to observe the agreement of theoretical delays with those determined by a trial in which the theoretical assumptions would be closely followed. This has now been performed at two trunk group sizes, $c = 2$ paths, loaded by approximately $a = 1.8$ erlangs or an occupancy of $\alpha = 0.90$, and $c = 10$ paths at an occupancy of approximately $\alpha = 0.80$.

For these throwdowns, random origination times were obtained through use of Tippett's Random Numbers. An hour was visualized as being composed of 100,000 (or, as in one case, 1 million) consecutive discrete intervals, numbered serially. Choosing 5 (or 6) digit random numbers then provided the start times of the subscribers' bids for service.

Likewise holding times were chosen by random numbers from an exponential universe by dividing it into 100 equal probability segments and assigning each a number from 00 to 99. A central value of holding time was chosen to represent the range of cases within each segment. The last segment, number 99, on the long tail was further subdivided into 100 parts in order to give more definition in the long call lengths which are believed to be critical.

A comparison of the proportion of traffic expected to suffer delays beyond various multiples of the average holding time as given by Rior-dan's theory for delayed calls served in random order, and by the throw-down results, is given in Figs. 1 and 2. As discussed below, the cases studied are considered to give satisfactory assurance as to the adequacy of the approximations involved in the theory.

The two trunk case based on 3000 calls submitted shows fairly good agreement with the theoretical distribution out to delays as large as 50 multiples of an average holding time which includes more than 99.5

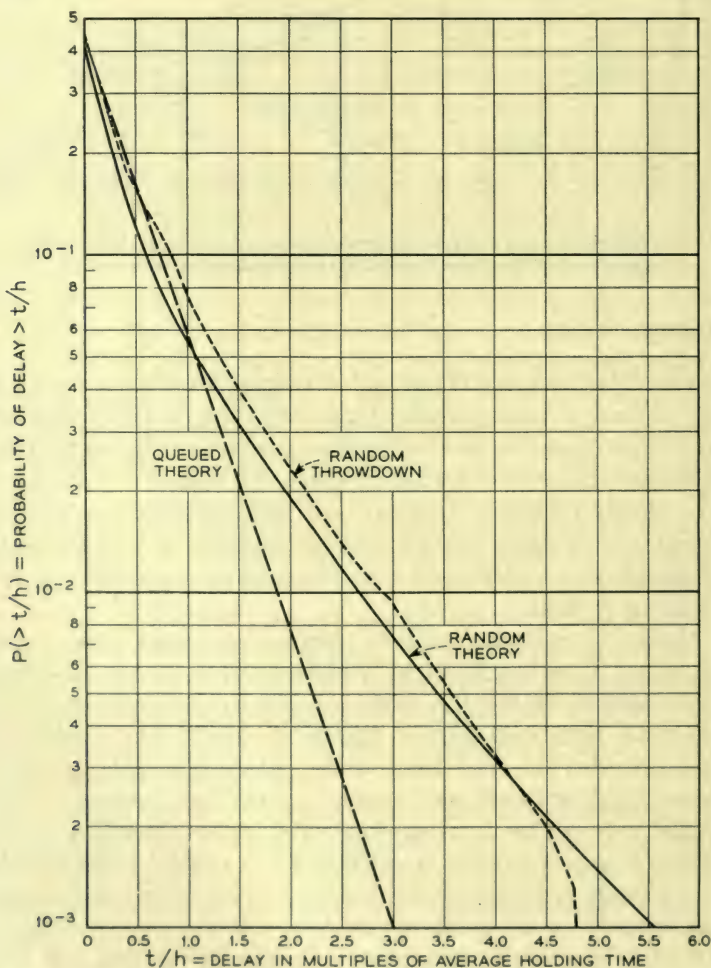


Fig. 2 — Distribution of delays. Theory versus throwdown, delayed calls handled at random, $c = 10$ paths, $\alpha = 0.80$, 1500 throwdown calls.

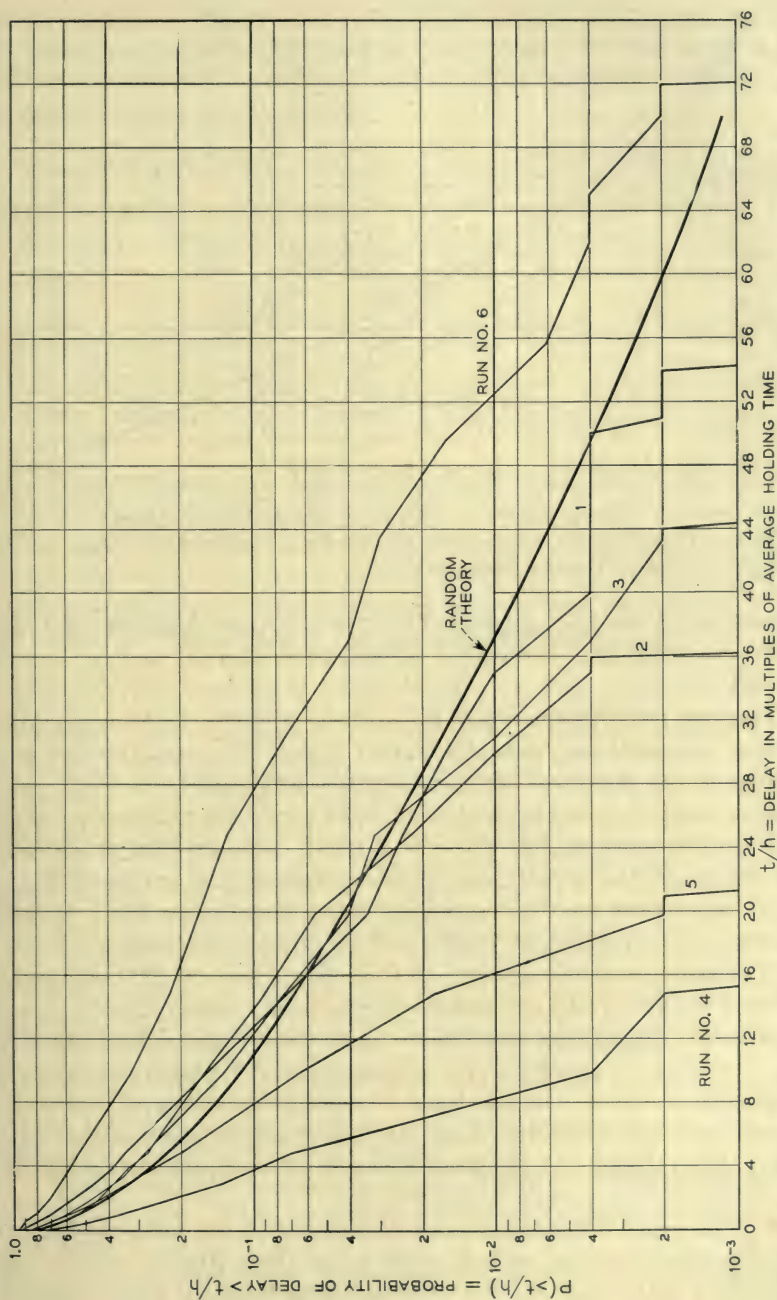


Fig. 3 — Distributions of delays on 2 paths by groups of 500 calls in random delay throwdowns, designed load $\alpha = 0.90$.

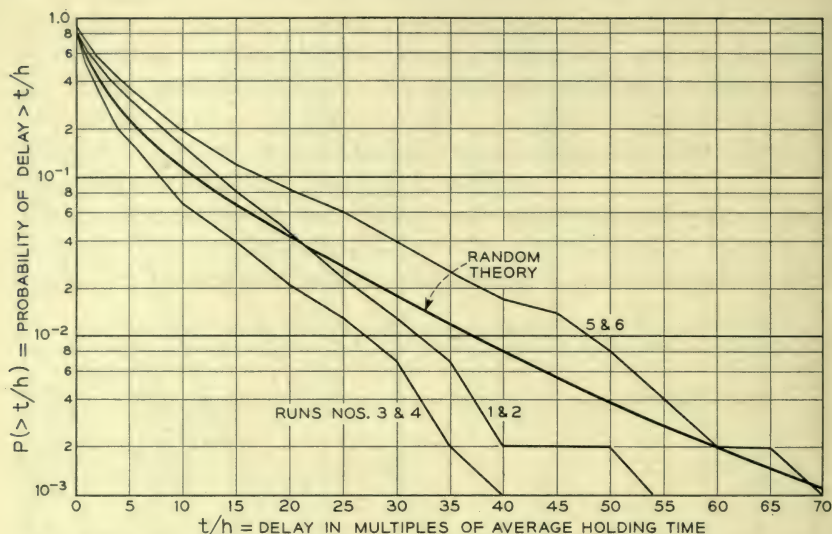


Fig. 4 — Distributions of delays on 2 paths by groups of 1000 calls in random delay throwdowns, designed load $\alpha = 0.90$.

per cent of all the calls delayed. The 10 trunk case based on 1500 calls at 0.8 occupancy also shows good agreement to 99.5 per cent of the calls delayed.

In making throwdown tests of this sort, the criterion for deciding when one has proceeded long enough is rather vague. The usual practice is to summarize the delays at regular intervals and observe at what point it seems likely that making additional tests would not change the results by a sensible amount. For the $c = 2$ trunk case, six runs of 500 calls each produced the several very different broken line curves of Fig. 3 shown superposed on the theoretical delay distribution for $\alpha = 0.90$. Clearly no one of these by itself could be given much weight.

Consecutive runs were paired to form three runs of 1000 calls each, as shown in Fig. 4. As one would expect, their spreads have narrowed appreciably. Combining these three runs yielded the dotted curve of Fig. 1, which, of course, has a correspondingly smaller likelihood of sampling error in it. On the basis of such a succession of narrowing spreads, one can, with some feeling of assurance, estimate within what narrow band about the observed curve the true unknown curve (approachable by many more tests) must lie.

On Figs. 1 and 2 the shapes and positioning of the total throwdown and theoretical curves seldom differ more than 20 per cent on the probability scale down to the $P = 0.005$ probability level. The dis-

parities measured along the delay axis in the higher ranges of the variable, are, of course, considerably less. A comparison of the theoretical and observed proportions of calls delayed, and the average delays on all calls is shown in the following table:

Trunk Group Size, c	Occupancy α	No. Calls in Throwdown	Proportion of Calls Delayed		Average Delay on All Calls in Multiples of Average Hold Time	
			Theory	Throwdown	Theory	Throwdown
2	0.9	3000	0.853	0.855	4.30	4.71
10	0.8	1500	0.409	0.444	0.205	0.254

These differences between theory and observation are well within the variations which would be expected with the lengths of throwdown runs made.

Further reassurance that the traffic submitted in the two throwdown tests originated in a manner reasonably similar to that assumed in the theory was obtained by making "switch counts" at regular intervals during the throwdowns from which frequency distributions, $f(x)$, of the number x of calls simultaneously present were constructed. These are shown in Figs. 5 and 6 for the two throwdown cases. The solid

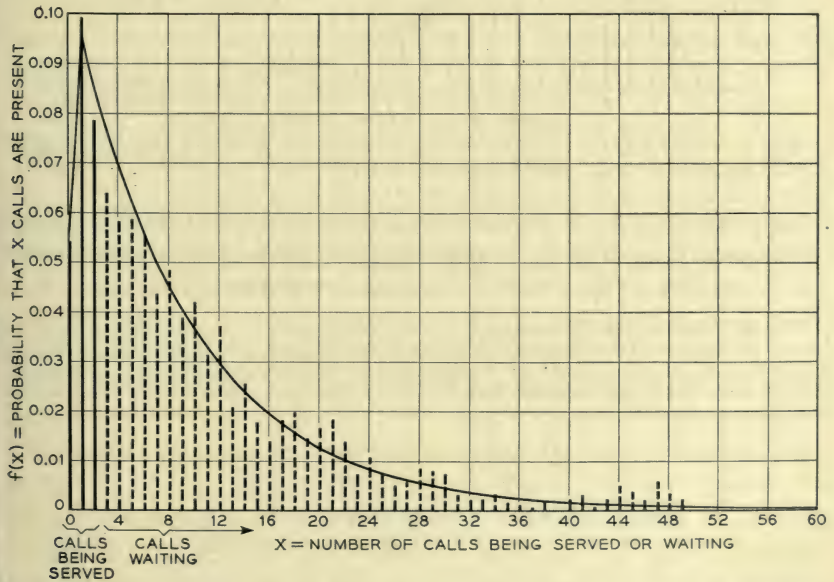


Fig. 5 — Distribution $f(x)$ of simultaneous calls. Theory versus throwdown, $c = 2$ paths, $\alpha = 0.90$, 3000 throwdown calls.

line spikes correspond to observations when all calls in the system were being served, that is $x \leq c$. The dotted spikes show those proportions of observations when one or more calls were waiting, that is $x > c$. The theoretical values of $f(x)$ are indicated by the smooth curves where they pass over discrete values of x . The theory and observations are seen to be in quite good agreement.

Referring again to the theoretical delays (and the throwdown checks) on Figs. 1 and 2, very much larger delays can obviously be obtained when delayed calls are handled at random than when they are handled in a strict first-come-first-served, or queued, order, the latter distri-

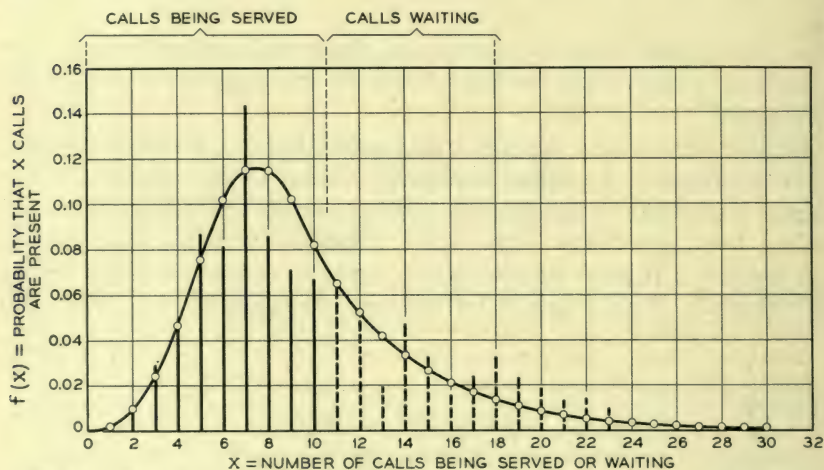


Fig. 6 — Distribution ($f(x)$) of simultaneous calls. Theory versus throwdown, $c = 10$ paths, $\alpha = 0.80$, 1500 throwdown calls.

butions being shown by the straight lines which start at nearly the same ordinates at delay 0 as the random handling curves, and cut down across the lower part of the charts.* Although fewer very short delays occur

* Delay curves for exponentially distributed holding time calls in systems where delayed calls are handled in order of arrival, are given by E. C. Molina in "Application of the Theory of Probability to Telephone Trunking Problems," Bell System Technical Journal, Vol. 6, p. 461, July, 1927. They are calculated from the Erlang equation

$$P(>t) = P(>0)e^{-(c-a)t} = \frac{\frac{a^c e^{-a}}{c!} \frac{c}{c-a}}{\sum_{x=0}^{c-1} \frac{a^x e^{-a}}{x!} + \frac{a^c e^{-a}}{c!} \frac{c}{c-a}} e^{-(c-a)t} \quad (1)$$

where the delay t is expressed in multiples of the average holding time. Values of $P(>0) = C(c, a)$ can be read approximately from Figure 21.

with this method of handling than when a random selection of the waiting calls is followed, the very long delays are markedly reduced, and on this account the queueing procedure is generally preferred. These effects are particularly evident at the higher occupancies. As illustrated in Fig. 1, the "queued" and "random" delay curves at an occupancy of $\alpha = 0.4$ show little difference down to the $P = 0.001$ delay level.

IMPERFECT QUEUEING

Interest has often centered in questions as to what form the delay curves might take in a system in which queueing of the calls is maintained to a limited extent, and beyond which the record of order of arrival would be lost. Such an instance might occur with a team of toll recording operators who were able to keep well in mind the order of arrival of signals up to a certain number waiting, whereupon they would lose track and not regain this ability until the number of waiting calls had again dropped below some small number. Other situations with actual or equivalent limited delay storage arrangements can readily be imagined.

To study a case of limited queueing, a short subsidiary throwdown was next run on the $c = 2$ case, using the 1000 calls of Runs 1 and 2 of Figs. 3 and 4 (which comprised the 1000-call sequence most closely approaching the theoretical distribution). Three rules for delayed call handling were tested:

- (1) Delayed calls are served in random order.
- (2) Delayed calls are queued (served in order of arrival).
- (3) Delayed calls are queued until more than w are waiting at which time their arrival order is lost and they are served at random. When the number waiting again drops below w , newly arriving calls are queued behind those randomized calls still waiting. Note that case 1 corresponds to $w = 0$, and case 2 to $w = \infty$.

The comparative results are shown on Fig. 7, with w given successively values of 0, 8, 20, 25, 30 and ∞ . The $w = 0$ curve, of course, is taken directly from Fig. 4 for Runs 1 and 2 combined. Although this curve does not agree particularly well with theory (Curve A), its movement with changes in w is nevertheless instructive. As seen, queueing as far as $w = 8$ waiting calls produced practically no improvement in the delay distributions. (Perhaps with the occurrence of such large numbers of waiting calls, reaching a maximum of 35, one could not expect queueing of so few as 8 to have much effect.) The next selection of $w = 20$, however, still showed only a relatively slight improvement, particularly in

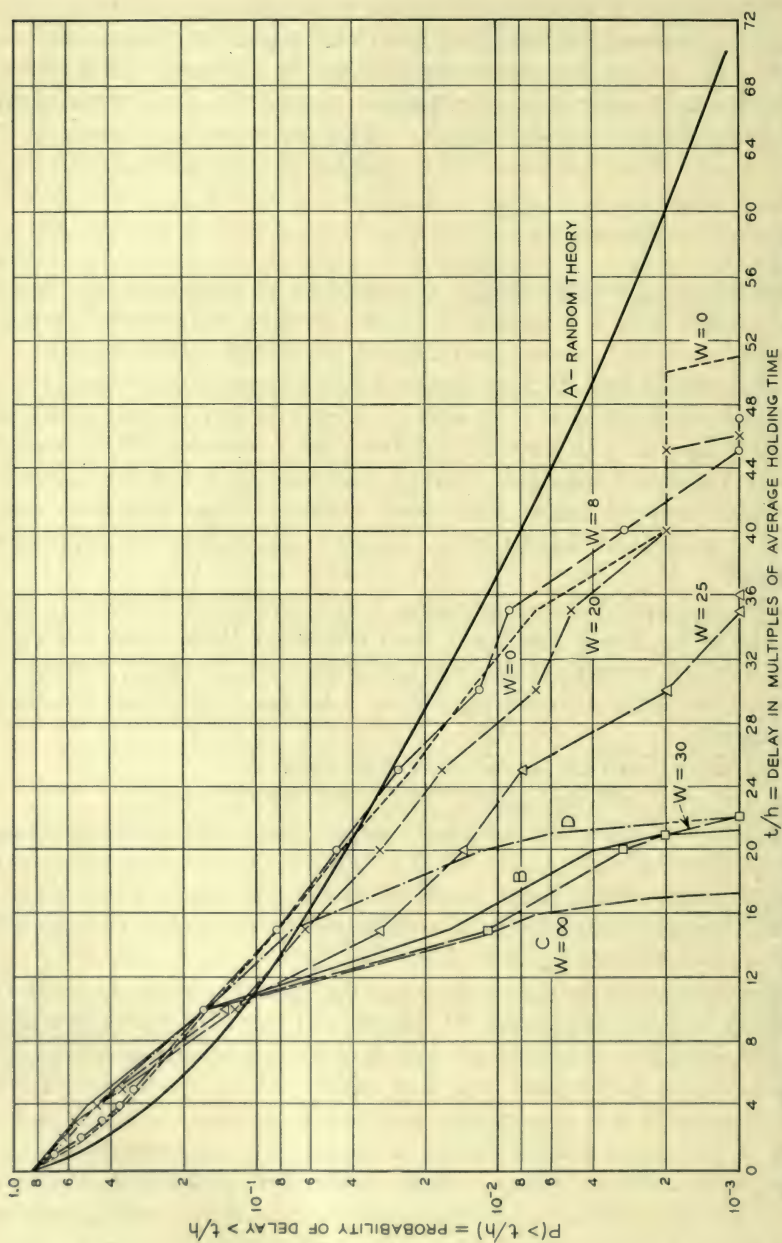


Fig. 7 — Throwdown delay distributions with imperfect queuing, $c = 2$ paths, $\alpha = 0.90$, 1000 throwdown calls.

the long delays occurring on a few of the more unfortunate calls. Even choosing $w = 25$ moved the delay curve hardly more than half way from the completely random to the fully queued curve of $w = \infty$. The $w = 30$ selection shows the accomplishment of nearly fully queued results, the latter being given by curve C ($w = \infty$). Thus one would apparently find little value in instructing a team of two operators working at an occupancy of 90 per cent to try to remember the order of arrival of waiting calls unless they could keep track of an unexpectedly large number.

Electrical storing circuits have long been used to assist the ordering of waiting calls. They have the especial advantage of not becoming confused and losing the order of the calls which have engaged them up to the limit of their storage capacity. In the Bell System two methods of approaching true queueing are in common use. In one method, such as found, for instance, in the No. 3 Information Desk, a number of storage circuits are provided so that as a waiting call is served from the number one storage position, all the others waiting on storage circuits drop down one position. If s such circuits are provided, and more than s calls have been waiting, one of the excess will then be chosen at random to occupy the newly vacated s th storage circuit.

The second method used widely in both local and toll systems is known as gating. In its simplest form a gate opens into a "corral" where the operators or other service media are located. So long as calls simultaneously demanding service do not exceed the number of operators (trunks, markers, etc.) the gate is ineffective. As soon as one call has to wait, the gate closes until that call obtains service, and then admits to the corral all calls which have accumulated on the outside. The gate again closes until all calls within the corral are served; and so on. Thus the calls are admitted in bunches to the corral. Between bunches there is strict queueing but within bunches when they get inside the gate the calls are substantially served at random. As long as the bunches are small the effect of true queueing is approached. In any event a strong safeguard against excessively long delays on a few unlucky calls is introduced. In the Bell System, a variety of gating plans are found such as double gates, gates with additional preferences for certain types of calls, and schemes for placing calls outside the gate again if they cannot be served immediately. Each of these must be studied with its own peculiar characteristics in mind.

To illustrate the effectiveness of the storage circuit type of automatic queueing arrangement, the 1000 calls of Runs 1 and 2, for the two path case, were processed by a throwdown through a two operator, 20 storage

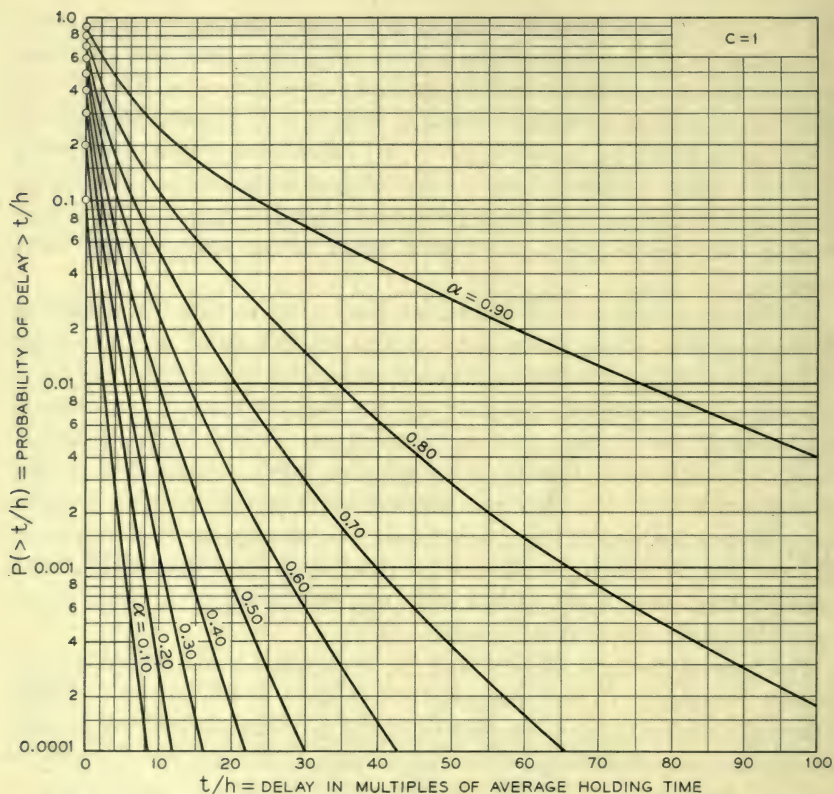


Fig. 8 — Delayed traffic served in random order, exponential holding times, $c = 1$.

circuit system. The resultant delay distribution is shown as Curve B on Fig. 7. (It is appreciated that this hardly represents a tolerable normal operating situation, but rather illustrates what the performance might be under extremely heavy traffic conditions.) The results are very close to those obtained with perfect queueing (Curve C) and show in striking fashion the gains in service to be made in certain delay situations by providing a limited storage apparatus with a memory not subject to confusion during moments of heavy overload.

When the 1000 calls of Runs 1 and 2 are submitted to the 2 paths through a simple gate in order to produce approximate queueing, the resultant delays are shown by Curve D on Fig. 7. Large improvements again occur in reducing the very long delays found with random handling. In fact by use of this simple (and usually relatively inexpensive) gating

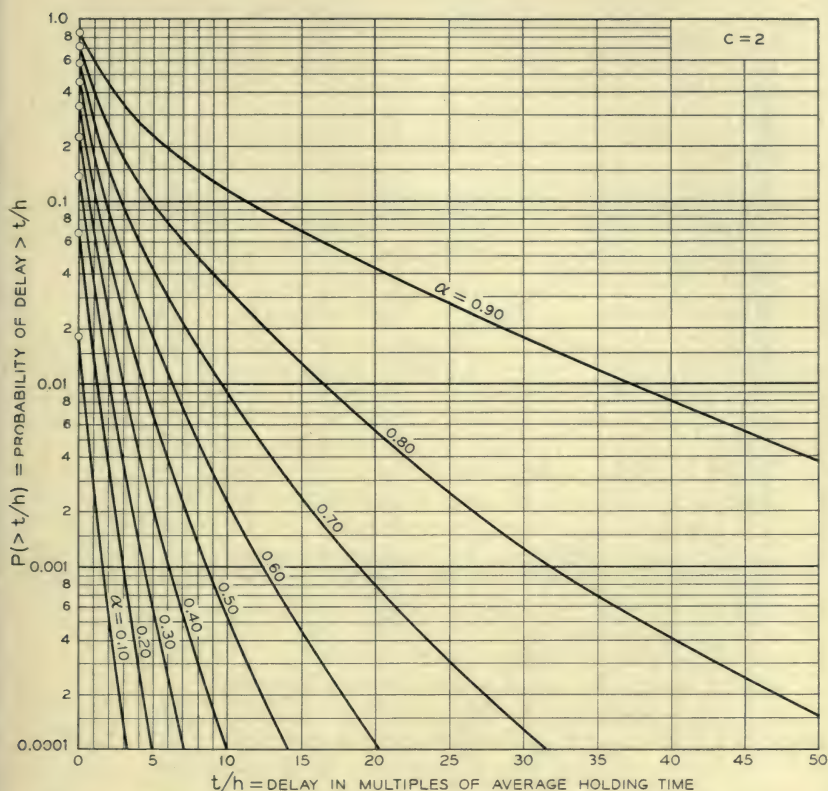


Fig. 9 — Delayed traffic served in random order, exponential holding times, $c = 2$.

scheme, delay results are obtained nearly as good as those realized by the provision of 20 storage circuits (Curve B).

WORKING CURVES

The adequacy of the Riordan theory when delayed exponential calls are served at random is believed to have been established and that it may be used with confidence to solve those practical problems where the underlying assumptions are well satisfied.

For working purposes, curves showing distributions of delays expected for occupancies up to $\alpha = 0.90$ and for group sizes of $c = 1, 2, 3, 4, 5, 6, 8, 10, 20, 50$ and 100, are shown in Figs. 8 to 18. These are plotted in the customary fashion with delay in multiples of average holding time

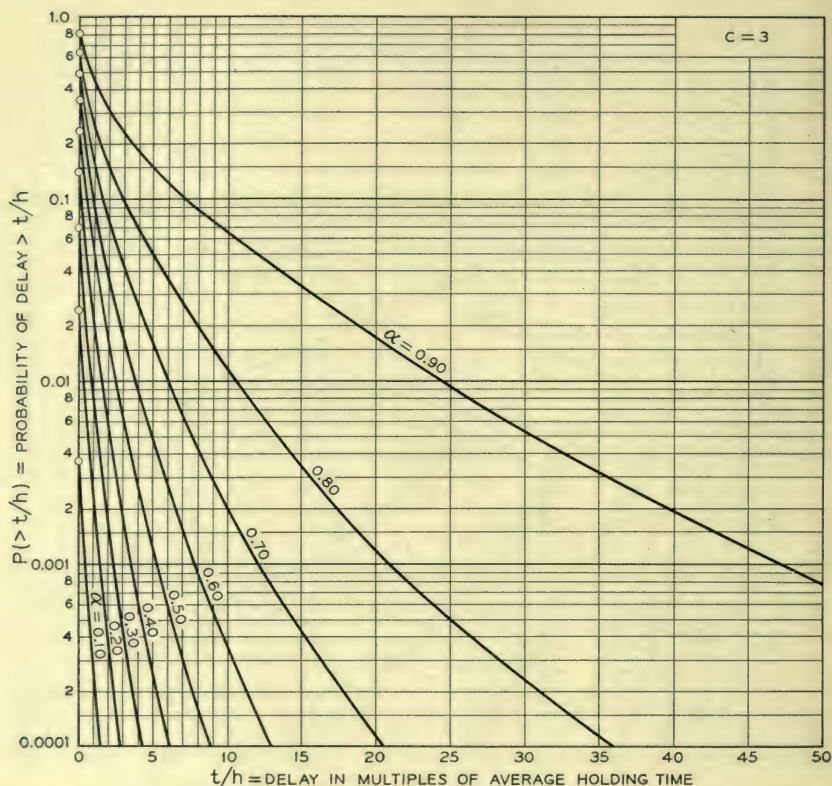


Fig. 10 — Delayed traffic served in random order, exponential holding times, $c = 3$.

as abscissa, and $P(>t/h)$, the probability of a random call meeting a delay greater than t/h , as ordinate.

Estimates of average delays, $\bar{t}$ (which are the same for queued and random service), are also commonly desired, and these are shown in Fig. 19. They are calculated from the equation

$$\bar{t}/h = P(>0)/(c - a) \quad (2)$$

If one wishes instead the average delay, $\bar{\bar{t}}$, on calls delayed, it may be obtained from

$$\bar{\bar{t}}/h = \frac{\bar{t}/h}{P(>0)} = \frac{1}{c - a} \quad (3)$$

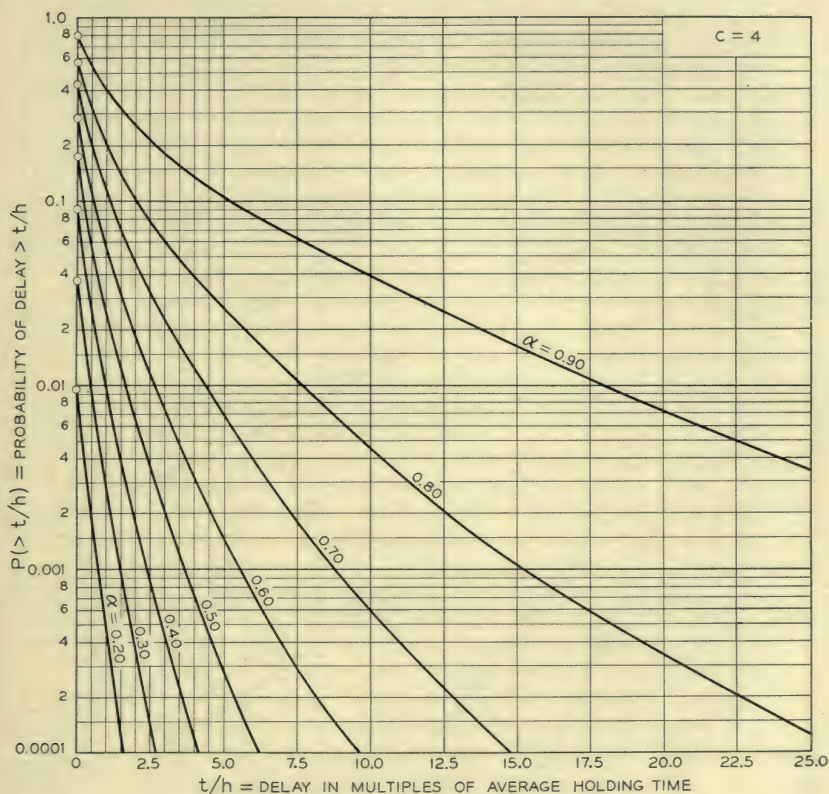


Fig. 11 — Delayed traffic served in random order, exponential holding times, $c = 4$.

ILLUSTRATIVE EXAMPLES

Example No. 1

A 20-trunk toll route carrying exponentially distributed holding time calls with average length of 5 minutes, is loaded with 16.0 erlangs of traffic, and any calls delayed will be served in random order. What per cent of all calls will be delayed? What per cent will be delayed more than 5 minutes? More than 10 minutes?

Solution. Enter Fig. 16 (the $c = 20$ chart) and read on the $\alpha = \frac{16}{20} = 0.80$ occupancy curve. At $t/h = 0$, the per cent of all calls delayed is found to be 26 per cent. At $t/h = 1$, the calls having delays exceeding 1 holding time, or 5 minutes, are 1.2 per cent, and at $t/h = 2$, the calls with delays exceeding 10 minutes are 0.2 per cent.

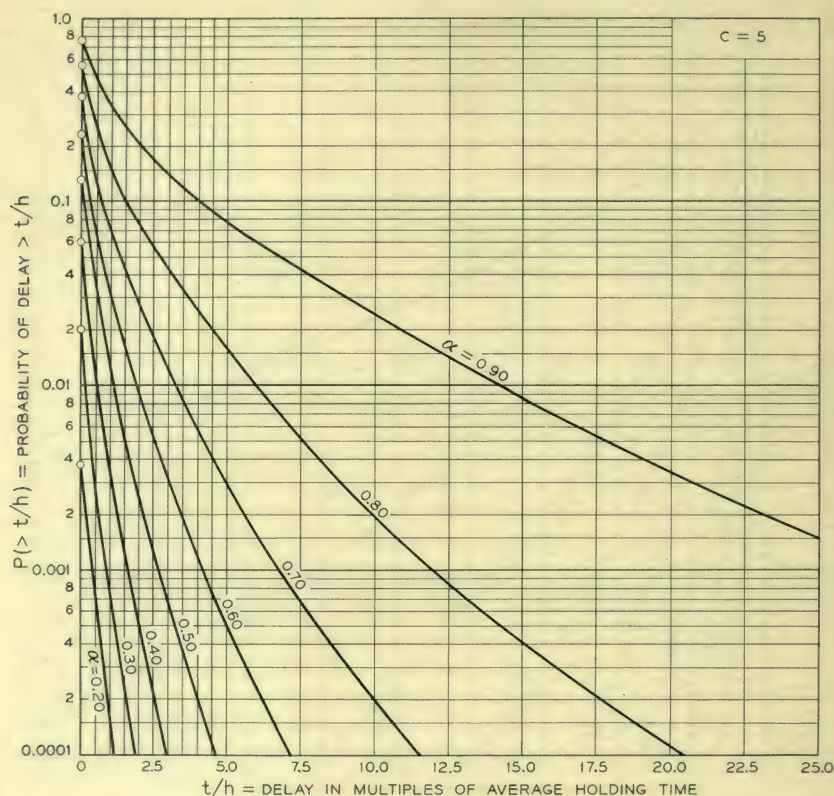


Fig. 12 — Delayed traffic served in random order, exponential holding times, $c = 5$.

Example No. 2

How many operators will be required at a department store telephone order desk to handle 225 calls per hour with an average delay not longer than half a minute, and with no more than 20 per cent of the calls delayed over 1 minute? Assume the average operator work time per call is 100 seconds, and waiting calls are handled in indiscriminate order.

Solution. The load to be carried is $a = (225)(100)/3600 = 6.25$ erlangs. The average delay, $\bar{t}$, is not to exceed $30/100 = 0.3$ holding time. Reading on Fig. 19, opposite an ordinate of 0.3 we select several trial values of trunks (operators) c , versus occupancy α , and form Table I, calculating the last column from the first two:

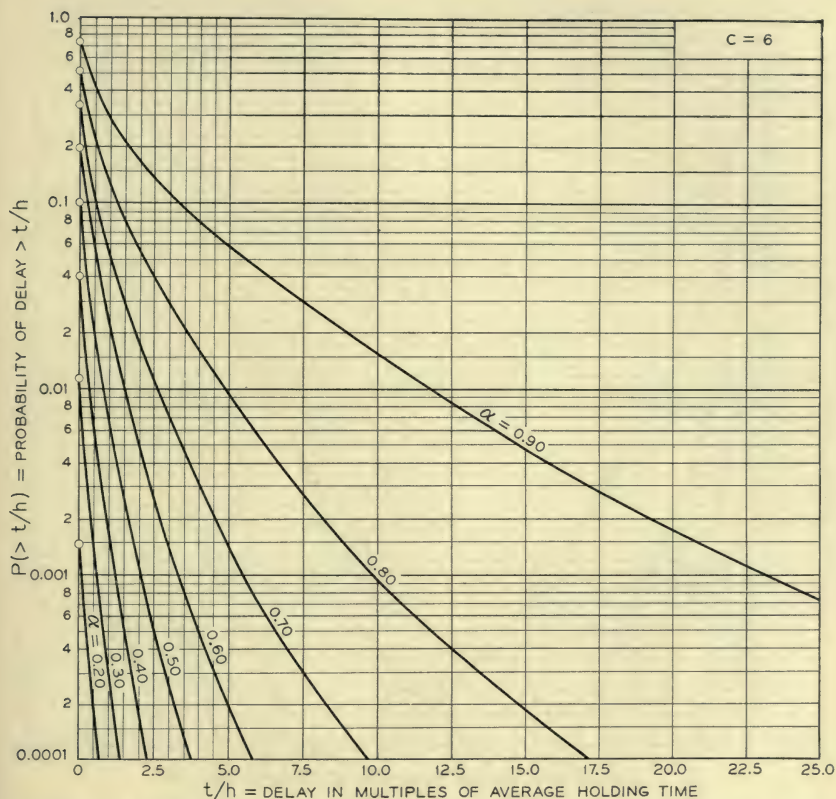


Fig. 13 — Delayed traffic served in random order, exponential holding times, $c = 6$.

To carry the 6.25 erlangs of traffic and meet the average delay requirement we see that 8 operators will be needed. Will 8 operators also fulfill the no more than 20 per cent delay over 1 minute requirement? Enter Fig. 14 (the $c = 8$ chart) with an occupancy of $\alpha = 6.25/8 = 0.78$. The per cent of calls exceeding a delay of $60/100 = 0.6$ holding time is about 12 per cent. A provision of 8 operators satisfies both requirements.

TABLE I

c	α	$a = c\alpha$
7	0.78	5.46
8	0.81	6.48
9	0.82	7.38

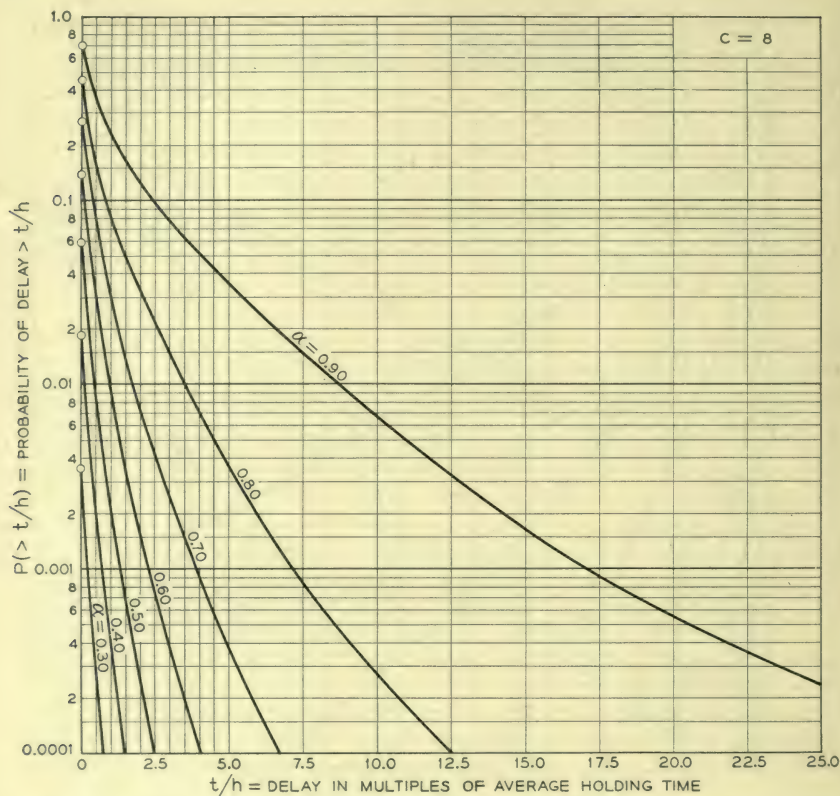


Fig. 14 — Delayed traffic served in random order, exponential holding times, $c = 8$.

Example No. 3

Suppose in Example 2, the second requirement had been that no more than one of 1000 customers should be required to wait over 3 minutes. Would 8 operators then suffice?

Solution. Reading on Fig. 14, with $\alpha = 0.78$ and $t/h = 180/100 = 1.8$, $P(>t/h) = 0.027$. Thus 27 in 1000 calls would be expected to experience delays over 3 minutes, and therefore more than 8 operators will be required. Consulting the $c = 10$ curves of Fig. 15, we find that with $\alpha = 0.625$, and $t/h = 1.8$, $P(>3 \text{ minutes delay}) = 0.0012$ which closely meets the one in a thousand requirement. Ten operators would then be needed; and this would, of course, (from Fig. 19) reduce the average

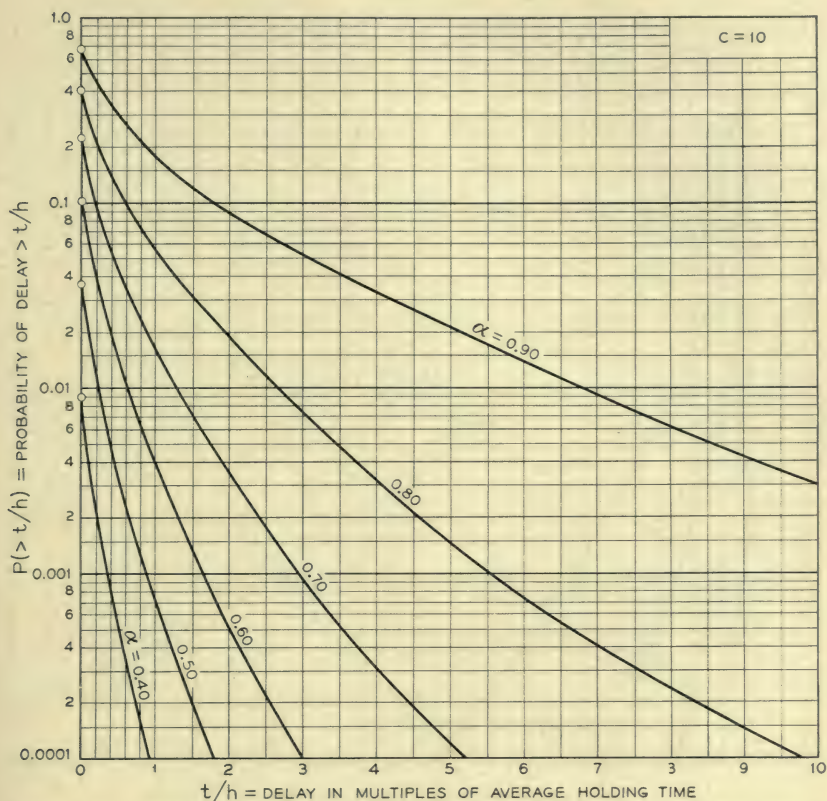


Fig. 15 — Delayed traffic served in random order, exponential holding times, $c = 10$.

delay on all calls to $0.035 (100) = 3.5$ seconds, an improvement in this characteristic of 7 to 1 over the 8 operator service.*

Example No. 4

How much improvement in the delay service would be obtained in Examples 2 and 3 by purchasing storing or gating equipment which would substantially insure calls being handled in order of arrival?

Solution. With 8 operators working at an occupancy of 0.78, the pro-

* Had some number of operators being required other than those for which working charts, Figs. 8 to 18, are supplied, intermediate values could be obtained by graphical interpolation, or better still by employing the basic Riordan chart, Fig. 20, combined with $P(>0)$ found on Fig. 21, to obtain delay versus load for any desired number of paths or facilities. This latter process is described in the Appendix.

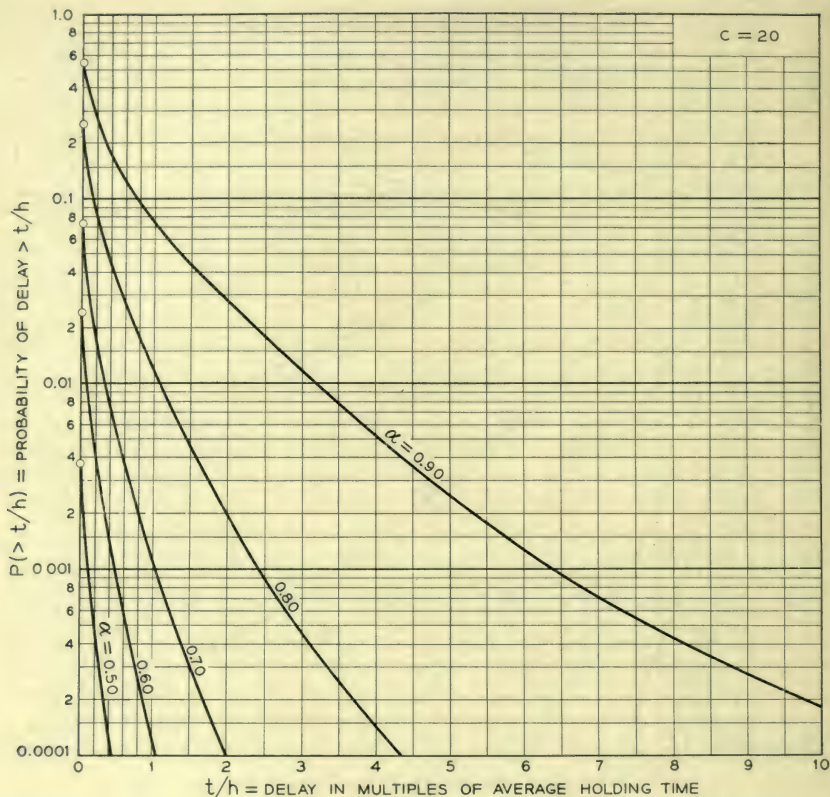


Fig. 16 — Delayed traffic served in random order, exponential holding times, $c = 20$.

portion of calls delayed is found to be $P(>0) = 0.41$ (Fig. 14). The probabilities of exceeding delays of $t/h = 0.6$ and 1.8 holding times are calculated for calls served in order of arrival by equation (1), in the following table:

t (Min.)	t/h	Queued $P(>0) = P(>0)e^{-(c-a)t/h}$	Random Handling
1	0.6	0.143	0.12
3	1.8	0.019	0.027

Comparing the queued and random handling of delayed calls one finds the perhaps unexpected result that with random handling some 2 per cent *fewer* calls are delayed longer than 1 minute than if perfect queueing had been present. This is due to the characteristic shapes of the two types

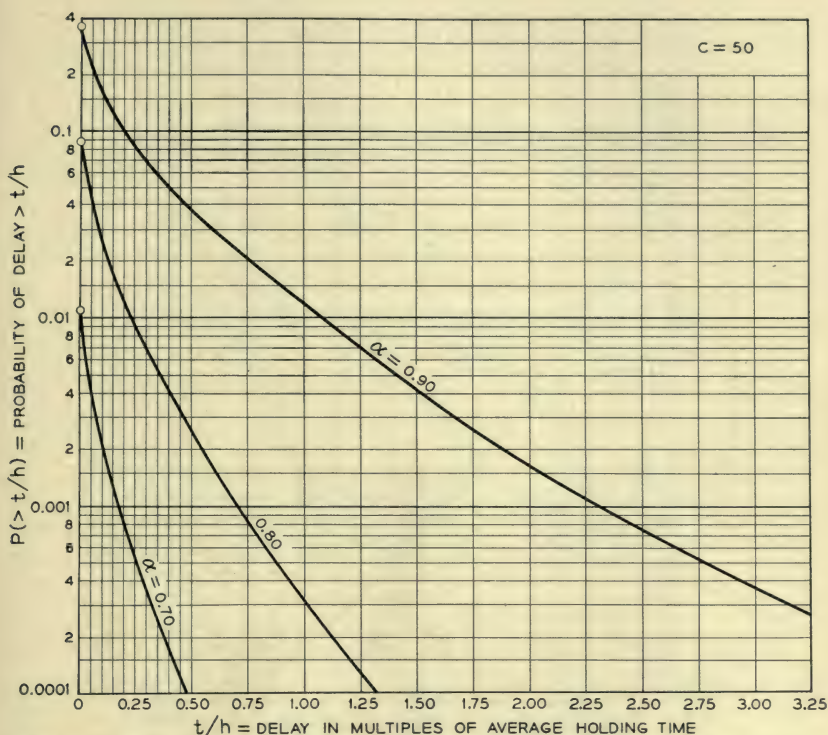


Fig. 17 — Delayed traffic served in random order, exponential holding times, $c = 50$.

of delay distributions, random handling producing more quite short and very long delays than does queueing. When a criterion of service is set at a relatively short delay, one may often expect it to be met more easily by not providing storing or gating circuits. On the other hand a criterion of service based on relatively long delays can nearly always be more readily met by the use of devices insuring partial or total queueing. In the example above the per cent of calls delayed longer than 3 minutes would be cut by a third through the use of queueing devices.

Example No. 5

Automobiles are parked in a large area adjacent to a State Fair grounds. There is one main exit through which two cars can pass at the same time. Upon leaving, drivers pay according to their parking time; and it requires, on the average, 20 seconds to complete the payment. If cars wish to leave during the afternoon busy period at a rate of 5.4 per

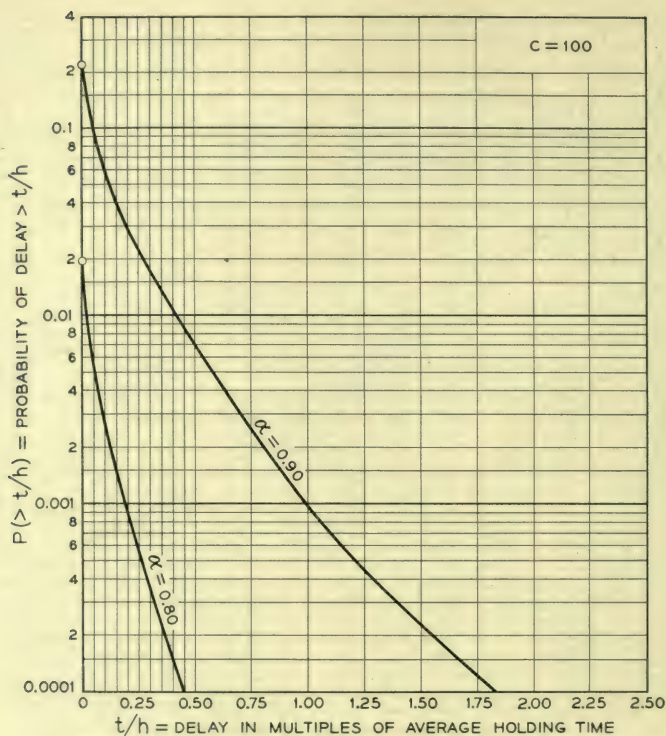


Fig. 18 — Delayed traffic served in random order, exponential holding times, $c = 100$.

minute, what per cent of the cars will be delayed more than 5 minutes? What will be the average delay for all cars?

Solution. Assume there is no traffic supervision and cars converge on the gate from many directions. Service in random order (or worse) among those delayed might then be approximated. Also the distribution of times for calculating and collecting the charge might be roughly exponential. We have then,

$$c = 2 \text{ paths}$$

$$\alpha = (5.4)(20)/(60)(2) = 0.90$$

$$t/h = \frac{5(60)}{20} = 15$$

Enter Fig. 9 at $t/h = 15$, read to the $\alpha = 0.90$ curve, opposite which find $P = 0.069$. Hence 7 per cent of the cars would be expected to have to wait 5 minutes or more. To obtain the average delay for all cars, enter

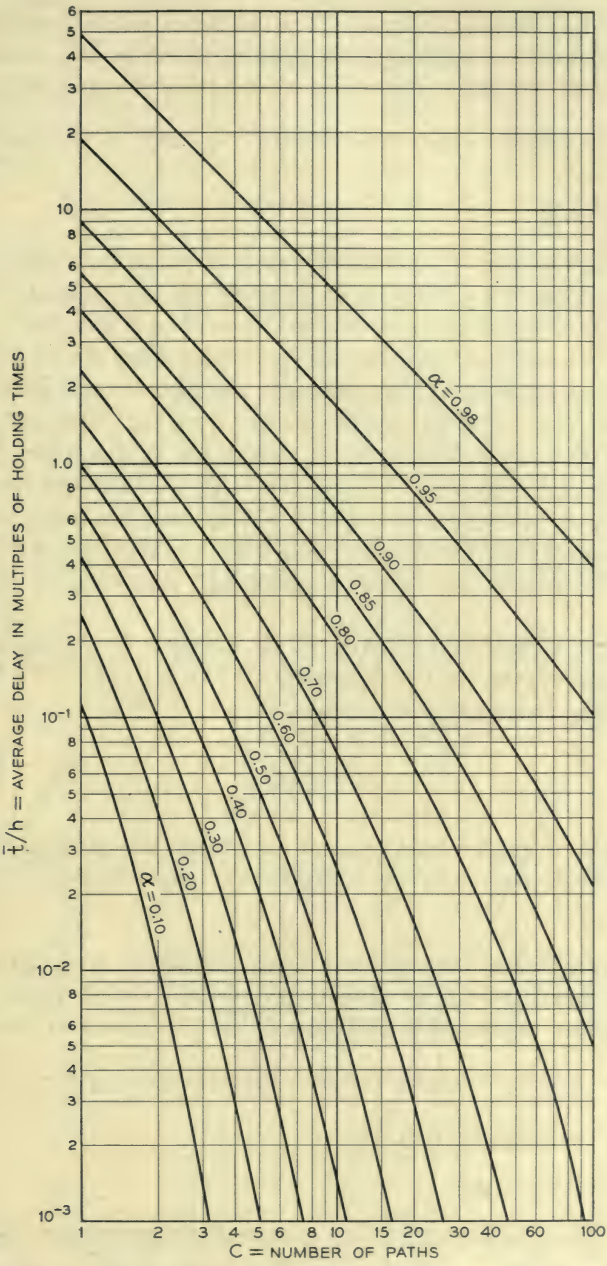


Fig. 19 — Average delay on all calls, exponential holding times.

Fig. 19 with the abscissa of 2 paths, read to the $\alpha = 0.90$ curve and find the average delay = 4.25 average holding times = 85 seconds. Or, one may obtain the same answer by substituting in equation (2),

$$\bar{t} = P(>0)h/(c - a) = (0.85)(20)/(2 - 1.80) = 85 \text{ seconds.}$$

Example No. 6

Suppose in Example 5, an efficient corps of police had been directing traffic toward the exit so that good queueing was maintained. What per cent of the cars would then be delayed more than 5 minutes?

Solution. We may now refer to other published delay curves for queued operation*, or, more generally, calculate the well known equation (1). In the present case we can read the answer from the "queued" curve of Fig. 1 as 4.2 per cent. Thus serving customers in the order of arrival nearly halves the occurrence of very long delays. (Note that the average delay for all cars remains unchanged at 85 seconds.) If a partial queueing were maintained the improvement would be intermediate, perhaps comparable with one of the "limited queueing" distributions shown on Fig. 7.

The author is indebted to Miss C. A. Lennon for constructing the working delay curves, and to Misses C. J. Durnan and J. C. McNulta for performing the throwdown checks.

APPENDIX

CALCULATION OF DELAY VALUES NOT FOUND ON THE WORKING CURVES OF FIGS. 8-18, FOR DELAYED EXPONENTIAL CALLS SERVED IN RANDOM ORDER

A master chart, Fig. 20, reproduced from Riordan†, gives in condensed form the *proportion* $F(u)$ of *delayed calls delayed* longer than u , where the delay is now expressed in multiples of the h/c ($u = ct/h$), and

c = number of paths (trunks, operators, etc.) provided

h = average holding time

t = delay time

To obtain the probability $P(>t/h)$ of any call being delayed longer than

* E. C. Molina, Ibid.

† Loc. cit.

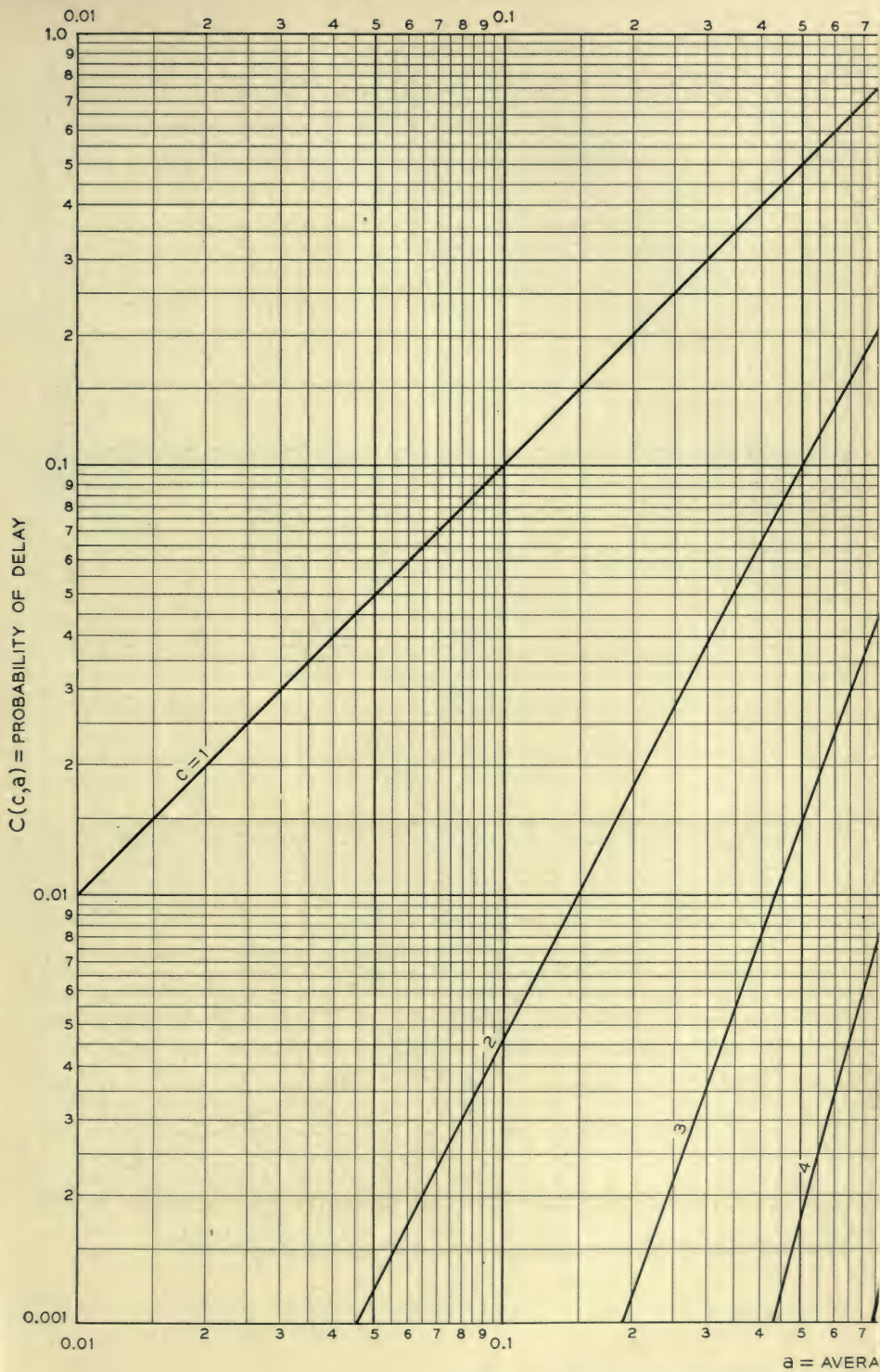
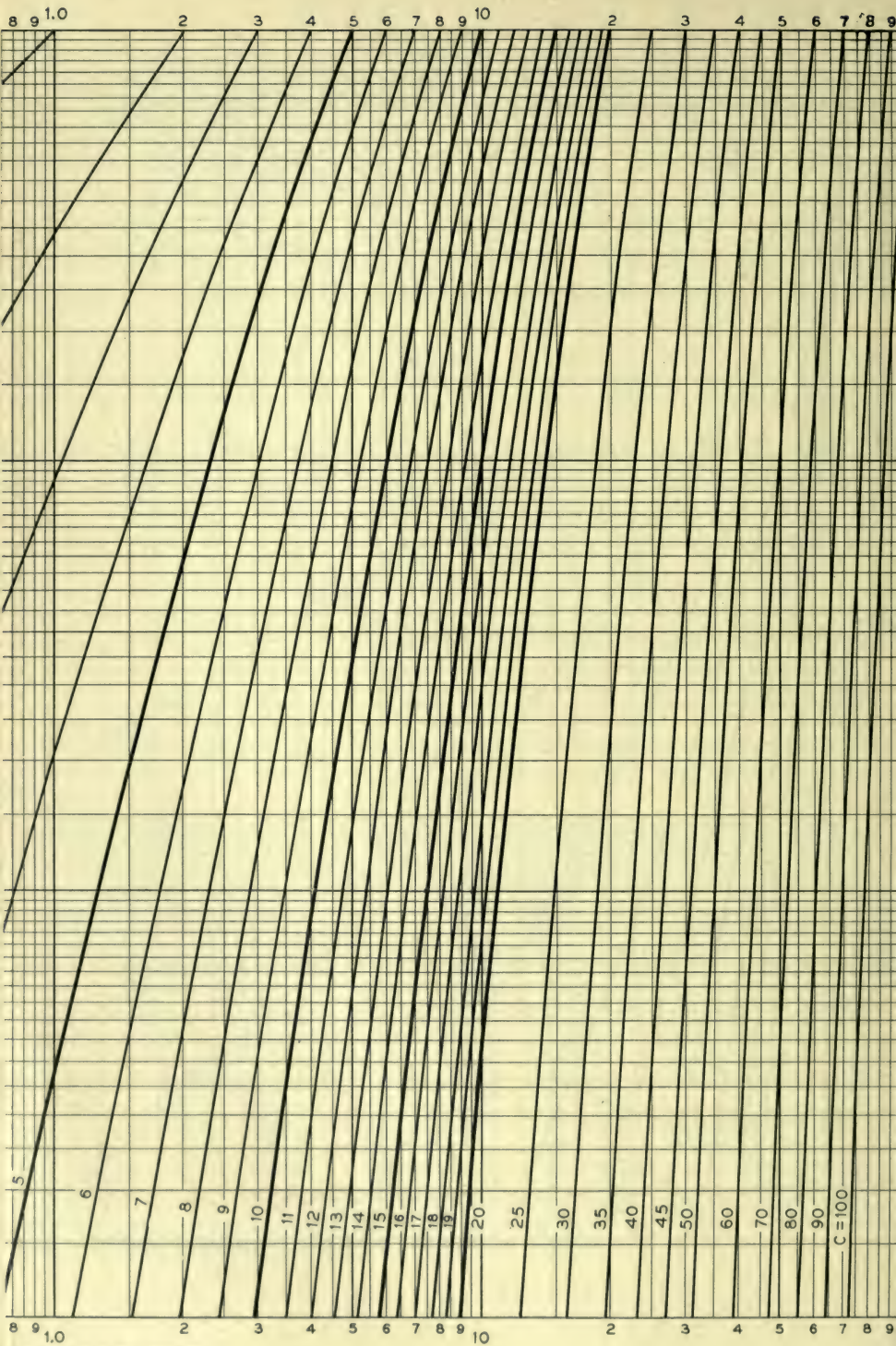


Fig. 21 — Probability of delay for exponen



GE LOAD SUBMITTED

tial holding time calls handled in a "delayed" basis.

t/h , we have

$$P(>t/h) = P(>0) F(u) = C(c, a) F(u) \quad (4)$$

Values of $P(>0) = C(c, a)$ are given for a wide range of a and c in Fig. 21. The application of equation (4) is quite simple.

Illustration 1. Suppose it is desired to obtain the probability of a call being delayed more than 3 holding times on a 10 trunk group without storage or gating circuits, and which carries $a = 9$ erlangs. Here $t/h = 3.0$, $c = 10$, $\alpha = 0.9$. Then $u = ct/h = 30$, and reading on Fig. 20 with this value of u , and $\alpha = 0.9$, we find $F(u) = 0.080$. Fig. 21 provides $C(c, a) = 0.67$ for $a = 9$ and $c = 10$. Substituting in equation (4),

$$P(>3 \text{ hold times}) = 0.67 (0.080) = 0.053,$$

which checks the value read directly from the $c = 10$ curves of Fig. 15.

Illustration 2. With an occupancy of $\alpha = 0.65$ on 15 paths what is the probability of meeting a delay greater than one holding time when delayed calls are served in random order? Calculate $u = ct/h = 15$. Enter with this abscissa on Fig. 20, and interpolating between the $\alpha = 0.6$ and 0.7 curves, read $F(u) = 0.022$. Fig. 21 shows for $a = 0.65(15) = 9.75$ and $c = 15$, $C(c, a) = 0.085$. Hence

$$P(>1 \text{ hold time}) = 0.085(0.022) = 0.0019.$$

Magnetic Resonance

PART II — MAGNETIC RESONANCE OF ELECTRONS

By KARL K. DARROW

(Manuscript received December 24, 1952)

Magnetic resonance of electrons is the analogue of magnetic resonance of nuclei, treated in the first part of this article. Though the analogy is close and the fundamental laws are identical, the two topics are remarkably different in detail. Though electrons are the commonest of particles, they display magnetic resonance only in somewhat exceptional cases. In many free atoms and most solid and liquid substances, magnetic resonance is suppressed by what is known as the "anti-parallel coupling" of electrons two by two. The exceptional cases are those of certain free atoms, ferromagnetic substances, and a restricted class of strongly paramagnetic substances; the resonance has also been observed very lately for the conduction electrons in metals. In the cases in which it does occur, resonance is likely to occur at a frequency or frequencies very different from that which the elementary theory predicts. This is sometimes because of the orbital motions of the electrons, oftener mainly because of the electric and magnetic fields existing in solids, and the deviations of the observed cases from the ideal case shed light upon these fields.

The subject of these pages is the magnetic resonance of electrons — "electron resonance" for short. Electrons being everywhere, one might expect it to be found in every substance; but for a fundamental reason it is a rare phenomenon, and this magnifies its interest. Those who search the literature for it under this its proper name will seldom find it, for it is frequently called "paramagnetic resonance" or, in appropriate cases, "ferromagnetic resonance." These are lengthy names which tend to veil the similarities between electron resonance and nuclear resonance, which latter was the theme of Part I of this article (in the January issue of this JOURNAL). I will introduce electron resonance by making use of all these similarities.

Magnetic resonance in general is due directly to the magnetism of

subatomic particles: nuclei and electrons. These, apart from the nuclei that are non-magnetic, may be visualized as minuscule barmagnets. The laws of resonance are determined by the fact that in a steady magnetic field, the magnetic moments of these particles may not point in any and every direction: instead, they are constrained to a finite and small number of what are called "permitted orientations." To each of these corresponds a special value of the energy of the little magnet in the field: thus the energy also is constrained to a finite and small number of "permitted" values. These are often called "Zeeman levels" or just "levels"; and the word "level" should be well known to those who are going to delve into the literature.

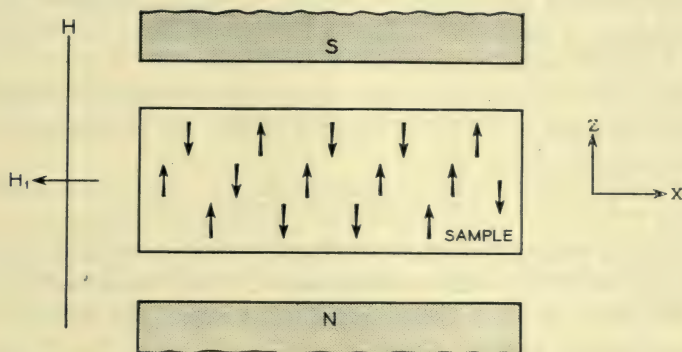


Fig. 1. — Scheme of the apparatus for observing magnetic resonance. The high-frequency circuits are omitted. The arrows within the sample may be taken as portraying the magnetic moments of either protons or electrons: their orientations are as given by the *old* quantum-theory.

Consider two orientations or levels of different energy-values. It will take *work* to turn the tiny magnet from the one of lesser energy to the one of greater energy. Magnetic resonance — and now I ought perhaps to speak specifically of magnetic resonance *absorption* — is such a turning. The agent of the turning and the source of the work is an alternating or oscillating magnetic field. The simplest cases are those in which the particle in question has only two permitted orientations. Many nuclei, among them the proton, belong to this class, and the electron belongs to it also. It is the analogy between proton and electron which I will develop.

Fig. 1 of this part is also Fig. 1 of Part I. The central rectangle depicts the *sample*, which for the study of proton resonance must be hydrogen or a compound thereof. The big arrow on the left represents a big magnetic field, of the order of several thousand gauss, which pervades the

sample; it is vertical and its strength is denoted by H . This is the field with respect to which the protons are oriented. These magnetic particles are represented by small arrows within the rectangle, the point of each arrow corresponding to the north pole of the corresponding proton. Slightly more than half of them are pointed in what I call the "up" orientation, which is that of lesser energy. The rest are pointed in the "down" orientation, that of greater energy. Magnetic resonance absorption of protons is the turning of "up" protons into the "down" direction.

The field which does the turning is an oscillating magnetic field with frequency (denoted by ν) in the radio-frequency range. It is horizontal, thus at right angles to the big field. It is produced either in a solenoid (the usual method for nuclear resonance) or in a resonant cavity (the usual scheme for electronic resonance) which encloses the sample but in Fig. 1 is left to the imagination of the reader.

Magnetic resonance occurs when the quantum-energy $h\nu$ of the oscillating field is equal to the work required to turn the proton from the up orientation to the down one:

$$h\nu = \text{work of turning} \quad (1)$$

h standing for Planck's constant. In Part I it was shown that the "work of turning" or energy-difference between the two orientations is equal to $2\mu_p H$: here μ_p stands for the magnetic moment of the proton, soon to be more carefully defined. Thus:

$$h\nu = 2\mu_p H \quad (2)$$

When ν and H are related by this equation one finds *proton resonance absorption*, which manifests itself by a splendid peak in the curve of absorption *versus* H for constant ν or the curve of absorption *versus* ν for constant H . For the frequency 42.6 megacycles the peak is found at $H = 10,000$ gauss.

To arrive at the basic formula for *electron resonance* we simply take (2) and substitute into it μ_e , the magnetic moment of the electron, for μ_p :

$$h\nu = 2\mu_e H \quad (3)$$

The magnetic moment of the electron is about 660 times that of the proton. Therefore if one works with such a field strength as brings the proton resonance into the radio frequency range, the electron resonance is to be sought in the microwave range. One might think that now I have said all that there is to be said about electron resonance; but this is only the beginning.

Much was said in Part I about the magnetic resonance of nuclei having more than two permitted orientations. We may seem to be wandering off the course if we revert to these, but this case is very pertinent.

There are nuclei with three, four, . . . up to ten or maybe more allowed orientations. One would expect them to display a multitude of peaks; but there is never more than one. This is for two reasons, which I give after introducing the symbol $(2I + 1)$ for the number of orientations. First, it is impossible to turn a nucleus from any orientation to any other *except* the nearest to the original one. This reduces the number of possible peaks to one fewer than the number of orientations. But second, all of these $2I$ possible peaks are of the same frequency for given H , or at the same field strength for given ν , so that they all coalesce into a single peak.

The formula for this apparent single peak which is strictly $2I$ coincident peaks has been derived in Part I, and this is it:

$$h\nu = (\mu/I)H \quad (4)$$

Now it is necessary to interpret I and μ ; and the interpretation is different according as one uses the old quantum theory or the new quantum mechanics. The old quantum theory deals more simply with these problems, and would be preferable if this field could be isolated from all the rest of physics; but the new quantum mechanics is worth the extra trouble that it causes.

In the old quantum theory, there are two definitions of I that reduce to the same thing. First, I is the angular momentum of the nucleus in terms of the unit $h/2\pi$; that is to say, the angular momentum of the nucleus is $Ih/2\pi$. Second, $Ih/2\pi$ is the maximum possible projection, upon the field-direction, of the angular momentum of the nucleus. This is because, among all of the allowed orientations of the nucleus, the one which is most nearly parallel to the field-direction is *exactly* parallel to the field-direction. So it was shown in Fig. 1.

In the new quantum mechanics, the second of these definitions remains valid and the first does not. This is because the orientation which is most nearly parallel to the field-direction is not exactly parallel thereto. It is inclined, in fact, to the field-direction by the angle $\arccos I/\sqrt{I(I+1)}$, and the angular momentum of the nucleus is $\sqrt{I(I+1)}(h/2\pi)$.

Thus there is one definition of I which is valid under both theories, and that is, that I is the maximum possible projection upon the field-direction, of the angular momentum of the nucleus in terms of the unit $h/2\pi$. Similarly it is always correct to say that μ is the maximum pos-

sible projection upon the field-direction, of the magnetic moment of the nucleus. But since these phrases are intolerably long, one avoids them by saying that I is the spin and μ the magnetic moment of the nucleus. In this sense, which to the users of quantum mechanics is a distorted one, the words "magnetic moment" shall be used hereafter.

The spin of the proton is $1/2$, and so is the spin of the electron. Equation (4) degenerates into (3) for the electron and into (2) for the proton. These we had already; what was then the point of introducing here the general case?

Well, the point is that two, or three, or several electrons may collaborate in what is known as "parallel coupling," though in the new quantum theory it is not quite parallel. They behave as though they formed a rigid unit, of which the spin is the sum of their spins and the magnetic moment is the sum of their magnetic moments. Thus if there are N of these electrons welded together (metaphorically speaking) it comes to the same as though there were a single particle of spin N times $1/2$ and magnetic moment N times μ_e . On putting these values of I and μ into equation (4) we find ourselves right back at equation (3), which is that for the individual electron. There is a single peak of magnetic resonance composed of N coinciding peaks, and it is just where the peak for a single electron would be. Thus in the ideal case, N electrons coupled parallel behave just like one electron by itself.

Such a conclusion may seem hardly worth the trouble of arriving at it; but note the stipulation "in the ideal case." This refers to what has been tacitly but obviously assumed till now, to wit, that no force acts upon the electronic magnet except the big field H . But there are also what I will call "local forces," forces due to fields within the sample arising from other particles in the sample. These forces may, and they often do, separate the N peaks which in the ideal case coincide. Often one finds a flock of resonance lines where, or near where, there should be only one; and if this is the explanation (which is not always the case, for there are other causes of "splitting") then the number of lines in the flock is the number of electrons coupled parallel.

This illustrates one of the great contrasts between the electronic resonance and the nuclear. Nuclear resonance is a "textbook phenomenon." The ideal case and the actual case are close together; the deviations due to the local fields are neither trivial nor useless, but they are not large enough to distort the simple laws, and it is quite permissible to leave them out of a first presentation. But the phenomenon of electronic resonance is liable to be distorted almost beyond recognition; and if one were to present only the cases in which the local fields are negli-

gible in effect, one's story would be relatively short and it would be grossly inadequate. But here the physicist, true to the tradition of his science, turns hindrance into help, and analyzes the distortions for the knowledge they are capable of giving about the fields prevailing in the sample. Thus whereas nuclear resonance is largely used for getting light on nuclei, the electronic resonance is largely studied for the information that it yields about the solid state.

Another of the great contrasts is due to what are called the "anti-parallel couplings" between electrons. Generally speaking (and this means: conceding an occasional exception) any type of nucleus of non-zero magnetic moment will display a detectable resonance if there are enough of them in the sample. Were this so with the electron, every substance whatsoever would display electron resonance. Experience shows that electron resonance is rare, usually conspicuous by its absence.

This is because electrons may, and not only may but usually do, pair off with one another in such a manner that the spin of such an "anti-parallel" pair is zero and so is the magnetic moment. There is no resonance for such a pair; and the customary absence of electron resonance signifies that in most solids, all the electrons are joined two by two into antiparallel pairs (this was known before magnetic resonance was first produced). I will call such electrons "compensated"; in this language, the substances in which magnetic resonance is to be sought for are those with uncompensated electrons. Mostly these belong to one or the other of two classes: the ferromagnetic bodies including the anti-ferromagnetic, and the "strongly paramagnetic salts." But there are a few other cases, and among these are those which are closest to the (unattainable) ideal of the perfectly free electron subjected.

THE NEARLY IDEAL CASES

Nearest of all to the ideal case are presumably the atoms which contain uncompensated electrons and are available for study by the molecular-beam method. Outstanding among these is the hydrogen atom, whose single electron must remain uncompensated because there is no other in the atom. About or quite as good are the atoms of sodium, potassium, and the other alkali metals, each of which contains a single uncompensated electron not to speak of several which are compensated. Moreover, these atoms are normally in a "ground state" in which the uncompensated electron has no orbital angular momentum. This hints at a complexity which is not always without influence on electron resonance, and must be mentioned here at the price of a detour.

Going back to ancient theory, let us imagine an electron revolving with frequency f in a circular orbit of radius r . It is equivalent to a current ef running continuously in the circular loop. According to the old theorem of Ampere, its magnetic moment is equal to the area of the circle multiplied by the current-strength; but the current-strength is to be expressed in *electromagnetic* units, so that the magnetic moment μ equals $(e/c)f\pi r^2$. The angular momentum p is mr times the speed of the electron, and therefore equals $2\pi mr^2 f$. For the ratio of the two we find:

$$\mu/p = e/2mc \quad (5)$$

This is what has lately been miscalled the "gyromagnetic ratio," a name which was originally applied and ought still to be applied to its reciprocal. It would be good to follow Gorter's suggestion of calling it the "magneto-gyric ratio."

I now state equation (5) in another fashion so as to introduce a symbol which is really a word, and is *the* technical word of this field of physics: it ought to be a word all spelled out, but it is just the letter g .

$$(\mu_{\text{orb}}/p_{\text{orb}}) = g(e/2mc), \quad g = 1 \quad (6)$$

Thus g is the ratio of magnetic moment to angular momentum given in terms of $e/2mc$ as unit, and *its value for the orbital motion of an electron is one*. Note also that though we have arrived at (6) in a very old-fashioned way, it is one of the results that have stood firm through all the mutations of quantum theory.

The study of what are known as "multiplets" in optical spectra led some thirty years ago to the conclusion that for the spin of the electron the magneto-gyric ratio is such that $g = 2$:

$$(\mu_{\text{spin}}/p_{\text{spin}}) = g(e/2mc), \quad g = 2 \quad (7)$$

This belief was substantiated by the "Dirac theory," and was not upset until measurements were made of the magnetic resonance of electrons in atoms by the molecular-beam method. The first such measurements were made upon atoms containing uncompensated electrons which had orbital motion as well as spin. I pass them over, and come direct to the most recent experiments on hydrogen atoms in their ground state, where there is no orbital motion of the electron to complicate matters. These are so recent that they came into print as these words were being written.

The hydrogen atom is a good example to take, not only for the reasons that I have given already, but also because it may be compared with the hydrogen molecule H_2 . The two electrons of the hydrogen *molecule* compensate one another, and there is no electron resonance. The two

nuclei — protons — of the molecule compensate one another in some of the molecules, enter into the parallel coupling in others. There are always some of these last in a beam of hydrogen molecules, and they produce the proton resonance of which so much was said in Part I. The *atoms* produce the electron resonance.

Look now again at equation (4), and remember that p is $Ih/2\pi$ — and remember that p is to be interpreted as the maximum permitted component, along the field-direction, of the angular momentum.

Consider now the experimenter with molecular beams of hydrogen *molecules* and hydrogen *atoms* at his disposal. In a magnetic field of field strength H he finds the proton resonance of the former at frequency ν_p , and ascertains (μ/I) of the proton by putting his data into equation (4):

$$(\mu/I)_p = h\nu_p/H \quad (8)$$

In the same field he finds the electron resonance of the latter at frequency ν_e , and ascertains (μ/I) of the electron similarly:

$$(\mu/I)_e = h\nu_e/H \quad (9)$$

Now he has both values; but the accuracy of both is contingent on the accuracy of the measurement of H , and this is not so good as he desires. However he can dispense with the measurement of H at a price — the price of getting his value of the magnetic moment of the electron in terms of units other than c.g.s. units. This is not a great sacrifice; Nature does not share our affection for c.g.s. units; there are others which are more suitable to the enterprises of the theorist.

If we divide (8) into (9) we get rid of both H and h . This means that if the experimenter measures ν_p and ν_e in one and the same applied field, he can evaluate (μ/I) for the electron in terms of (μ/I) for the proton without bothering about the values of H and h . Since I is the same for both particles, he obtains the ratio of the magnetic moments of electron and proton. The value of this ratio would be precious in itself, even if one had not the faintest idea of the value of either moment in c.g.s. units. It is 658.2288 ± 0.0006 .

It is also feasible to get the value of (μ/I) for the electron in terms of the "unit" $eh/4\pi mc$. This entity is so important that it has a name of its own: it is called "the Bohr magneton."

There is also a combination of experiments by which $(\mu/I)_e$ may be evaluated in terms of the unit $(eh/4\pi mc)$. This unit is so important that it has a name of its own: it is called "the Bohr magneton." The reader can easily show for himself that (μ/I) in terms of this unit is none other than the quantity g , of which this is a second definition (not identical with that of g in Part I).

The frequency of the proton-resonance, ν_p , is compared in a special experiment with what is known as the "cyclotron frequency," ν_c , of the electron. A free electron, projected at right angles to a magnetic field H , describes a circle in the plane perpendicular to the field. The frequency with which it makes the tour of this circular orbit is given by the equation:*

$$\nu_c = 2(He/4\pi mc) \quad (10)$$

If this frequency is determined in the same field as has served or is to serve for the location of the proton-resonance, we have:

$$(\mu/I)_p = 2(eh/4\pi mc)(\nu_p/\nu_c) \quad (11)$$

and consequently:

$$(\mu/I)_e = 2(\nu_e/\nu_p)(eh/4\pi mc)(\nu_p/\nu_c) \quad (12)$$

So here is the value of (μ/I) for the electron expressed in terms of the Bohr magneton, determinable by measurements on ratios of frequencies only! At this point the reader may well wonder why I did not eliminate ν_p from (12) by simply dividing it out. The reason is that one group of experimenters has determined (ν_e/ν_p) at one fieldstrength and another group of experimenters has determined (ν_p/ν_c) at another fieldstrength, so that ν_p does not have the same value in the two brackets: this is trivial.

The old belief, as I remarked above, was that (μ/I) for the electron amounted to exactly two Bohr magnetons. But the combination of two experiments which I have just so sketchily described has led to the following result for the electron in the hydrogen atom:

$$(\mu/I)_e = g(eh/4\pi mc), \quad g = 2.002292 \pm 0.000024 \quad (13a)$$

But is this truly the ideal case? Defining the "ideal case" as that of the free electron, remembering that the electron in the hydrogen atom is bound even though lightly bound, and making what is deemed the appropriate correction, one elevates the foregoing value of g by 35 parts in a million, and obtains:

$$\text{Ideal } (\mu/I)_e = g(eh/4\pi mc), \quad g = 2.002327 \pm 0.000024 \quad (13b)$$

Thus the old belief was wrong by about one part in a thousand. Be it mentioned in passing that the Dirac theory which led to $g = 2$ has been modified in the meantime by what is known as "quantum electrodynamics", which gives a good account of this result.

* To be derived by equating the force $Hv(e/c)$ exerted by the field upon the electron to the "centrifugal force" mv^2/r ; here v stands for the speed of the electron and r for the radius of the circle.

Since I is $\frac{1}{2}$ for the electron (as it is for the proton) the magnetic moment of the free electron is:

$$\mu_e = (\frac{1}{2})g(e/4\pi mc) = 1.001146 \pm 0.000012 \text{ Bohr magnetons (14)}$$

This is the value which is 658.2288 times the moment of the proton.

Another case very near to the ideal is afforded by the electrons of such atoms as manganese widely dispersed in a phosphorescent solid. Thus, there exists a measurement of g made upon "zinc sulphide phosphor" containing manganese atoms in a concentration of 0.001 per cent. The value is 2.0024 ± 0.0004 . It must be said that the resonance in question is complicated both by fine structure and by hyperfine structure, terms to be explained in following sections. It is therefore necessary to use theory to locate, among the complex of peaks, the frequency which corresponds to the appropriate value of g .

Still another case which is close to the ideal is provided by the "F-centres" in colored crystals, mention of which was made in Part I. An F-centre is a cavity in a crystal lattice occupied by a free electron batting around, as I said in Part I, like a wild animal in a cage. Several physicists have found their resonance, present when the crystal is colored and absent when the crystal is bleached. One, who produced the coloration by neutron-bombardment, located the peak at $g = 2.00$. Others report 1.995 ± 0.001 .

Still another case which is close to the ideal is afforded by the conduction electrons in a metal. These are so numerous that one might expect that the electron resonance that they produce must be extremely prominent. Yet the first such peak to be observed has been reported only as these lines are being written! The reasons for its inconspicuous character are two: most of the conduction-electrons are coupled anti-parallel, and the skin-effect confines the oscillating field in a conductor to a very narrow region close up against the surface. The second of these hindrances is overcome by using a colloidal dispersion of the metal, of which the spherules are less than 10^{-3} cm in diameter. Data are available (though not yet all in print) for lithium, sodium and potassium. The values of g are within a few promille of 2.000; the differences between these and the "ideal" value are small but not trivial, and in the case of lithium have been explained.

ELECTRON RESONANCE IN PARAMAGNETIC SOLIDS

There are paramagnetic solids that display the electron resonance. A magnificent illustration is shown in Fig. 2, belonging to an organic

substance of which the name and the structural diagram are included in the figure. This is one of the strongest and sharpest electron-resonance peaks on record. The g -value is 2.0064 ± 0.0002 ; it is therefore *almost* an ideal case, but the difference from the ideal value is sure and significant. It would however be misleading to suggest that such a case is typical.

What are called the "strongly paramagnetic salts" form a group with several features in common. They tend to have long names, and they have complex chemical formulae; crystal lattices or at any rate unit-cells which are non-cubic; and atoms some of which belong either to the rare-earth elements or to the elements of the "first transition group," iron or cobalt for instance. These atoms are likely to have two or more uncompensated electrons in parallel coupling. I now recall what was said about such coupled electrons in the introductory passage.

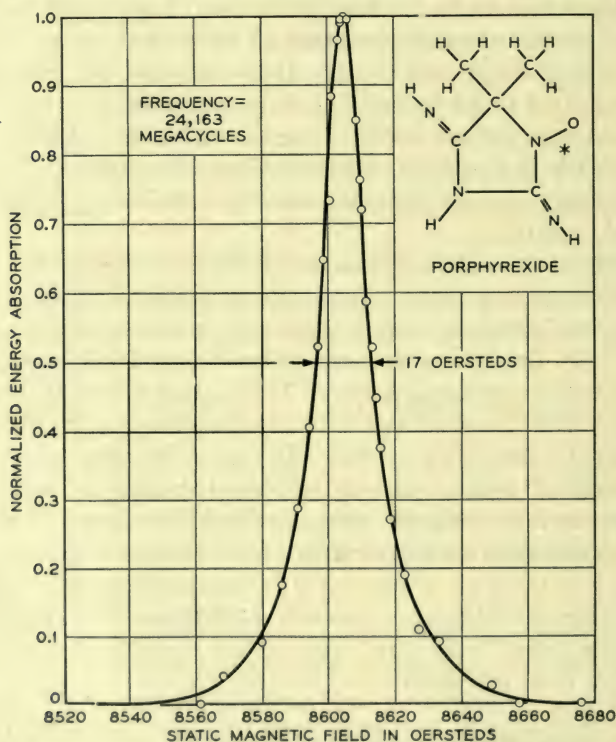


Fig. 2. — Electron resonance of porphyraxide. This is one of the strongest and sharpest peaks of electron-resonance yet observed. The g -value is 2.0064 ± 0.0002 , which makes it slightly but significantly different from the ideal case. In the structural diagram, the asterisk signifies a three-electron bond. (A. N. Holden, W. A. Yaeger and F. R. Merritt).

Two or more electrons — N electrons, let me say — may form, in effect, a rigid unit having a total spin $S = N/2$ and a total magnetic moment $N\mu_e$. Such a unit will have $(N + 1)$ allowed orientations in the big magnetic field. These will engender N resonance-peaks. In the ideal case, all of these would have the same frequency $2\mu_e H/h$, and would therefore coalesce into a single peak at the position appropriate to $g = 2.0023$. But in these crystals we are likely to find cases far from ideal, because of the conjoined influence of two factors. These are the presence of *orbital* motions of the electrons, and the presence of a big electric field within the crystal.

Were the atoms in question free, we could allow for the orbital motions. There would be a single resonance-peak, corresponding to a value of g which could be computed by a formula well known and much used in optical spectroscopy. Incidentally, this formula was used in interpreting the earliest molecular-beam experiments (not here described) that were the first to show that g in the ideal case is not exactly equal to 2.

Now, however, we are dealing with resonating electrons that are in a strong electric field, and moreover, an electric field which is usually unsymmetrical. If the asymmetry is sufficiently great, the orbital motions suffer a singular effect. This effect is known as “quenching.” It is impossible to explain and difficult even to describe without invoking quantum mechanics. One may say that the orbital angular momentum is no longer constant in time, and the associated magnetic moment almost but not quite disappears.

The spin survives the quenching; but it would not be right to say that the quenching restores the ideal case. The resonance is affected by what have been called the “remains” of the orbital magnetic moment. These have the following consequences:

(a) The N resonance-peaks, which coincide in the ideal case, may be drawn apart. They then form a group of N separate peaks, which is known as a “fine-structure pattern.” The number N tells us the number of electrons coupled parallel in the atom, for these two numbers are the same. Often the number of electrons coupled parallel is known from independent evidence, and in such cases it is confirmed by the number of lines in the fine-structure pattern. Sometimes it is not otherwise known, and in such cases it is identified with the number N .

(b) The value of g corresponding to the centre of the fine-structure pattern may be altered considerably from 2.0023, falling as low as 1.35 or rising as high as 6.5. This is as though a part of the orbital magnetic moment were added to or subtracted from the magnetic moment of the spin.

(c) The value of g may depend upon the orientation of the applied magnetic field with respect to the crystal.

(d) The frequency of the resonance-peak or peaks may not be proportional to H . In fact, it may deviate so far from being proportional to H that extrapolation to $H = 0$ will indicate that even in the absence of an applied magnetic field there would be a separation of the levels. Thus the asymmetric electric field within a strongly paramagnetic crystal may by itself produce the effect, which hitherto we have been ascribing entirely to the applied magnetic field. This is called "zero-field splitting."

One sees only too well that the interior of a strongly paramagnetic salt is no place to look for the ideal case, and that resonance in such a salt is a theme for deep study and not for facile interpretation. As a matter of fact, electron resonance in paramagnetic salts is valued for its contribution to our knowledge of the electric fields in these crystals; which is to say, that it is a part of solid-state physics, the details of which lie beyond the scope of this article.

HYPERFINE STRUCTURE OF ELECTRON RESONANCE

One of the most beautiful phenomena in this province of physics — and, I venture to say, not in this province only but in the whole of physics — is the "hyperfine structure" or "hyperfine splitting" of the electronic resonance. Here we see the spin and the magnetic moment of the nucleus collaborating with those of the electron to produce an exquisite and lucid joint effect. It is still the *electronic* resonance, and must never be confused with the nuclear resonance; but the single resonance-peak of the ideal case is split into a group of peaks, the number of which is determined by the spin of the nucleus.

Fig. 3 relates to neodymium — not however to the metal, but to neodymium atoms in a salt of neodymium, diluted with a salt of another metal so that the neodymium atoms may not influence one another through undue proximity. Neodymium is an element with two "odd" isotopes — that is to say, isotopes of odd mass-number — and several "even" isotopes. The even isotopes have non-magnetic nuclei, and so do not perturb the electron resonance. Each of the two odd isotopes has a nucleus of spin $7/2$ and non-zero magnetic moment. Such a nucleus will have eight permitted orientations in the big magnetic field. It will produce a local magnetic field in the region of the resonating electrons, and the strength of this field will depend on the orientation. The resonance-frequency depends on the big field compounded with the local field (we met with instances of this rule in the study of nuclear reso-

nance). Therefore there are eight resonance-peaks for the electrons in the atoms of the isotope 143, and eight more for the electrons in the atoms of the isotope 145. This is the key to the remarkable pattern shown in the curve at the bottom of Fig. 3.

In the middle of the pattern is the stump of a tall peak. This is the unperturbed peak due to the electrons in the atoms of even isotopes, those of which the nuclei have no magnetic moment. Whether it is at the position corresponding to $g = 2.00$ will depend on whether the displacement due to electric fields in the crystalline salt of neodymium, with which these data were obtained, is negligible or is not. Then, there are eight much shorter peaks. These are due to the electrons in the atoms

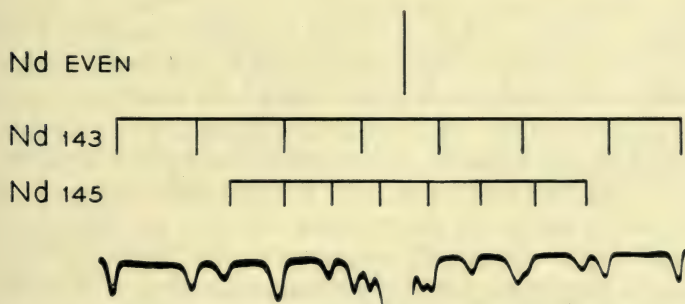


Fig. 3. — Hyperfine-structure pattern of the electron resonance of neodymium in a salt of the metal, showing that the nuclei of each of the odd isotopes of neodymium have eight orientations and therefore a spin of $7/2$, and that the even isotopes do not affect the resonance. (Courtesy of B. Bleaney).

of the more abundant of the two odd isotopes. Then, there are eight still shorter peaks (provided we count one which is merged with one of the other group of eight). These are due to the electrons in the atoms of the less abundant of the two odd isotopes. This is beautifully confirmed by the fact that the statures of the two groups of peaks stand to one another in the ratio of the abundances of the two isotopes! Further, the spacings within the two groups stand to one another in the ratio of the magnetic moments of the nuclei of the two isotopes. As for the two combs that stand above the curves, they are markers to identify for the onlooker the members of the two groups of peaks.

Observations on the similar pattern of a (rare) isotope of vanadium — vanadium 50 — have led to the inference that this nucleus possesses a non-zero magnetic moment and a spin equal to 6 (the highest value so far known). This may seem surprising, since I have implied that nuclei of even mass-number have neither spin nor magnetic moment. Vanadium

50 is however a nucleus with an *odd* number of protons and an *odd* number of neutrons. Such nuclei, of which there are only a few stable examples, (in Part I we met with two, the deuteron and N^{14}), are not bound by the usual rule.

FERROMAGNETIC RESONANCE

Ferromagnetic bodies owe their distinctive feature to uncompensated electrons. This suggests that the magnetic resonance of electrons will be discernible in such bodies, and so indeed it is. In this case it is commonly known as "ferromagnetic resonance." However, unless the sample is in the shape of a sphere, the resonance-peak will be found in what appears to be very much the wrong place. This is due to the magnetization of the substance, which produces a remarkable effect upon the location of the resonance. The field strength in the region occupied by the sample, which would be H if the sample were not there, is changed to a very different value; and yet in general it would not be right to take the value of H_i , the "internal field strength" and put it in place of H in equation (3). We must understand this effect and make the proper allowance for it before we do anything else with the data (unless, I repeat, we confine ourselves to data obtained with spheres). The effect appears to be beyond the power of "intuition" to conceive, and we must have recourse to the fundamental equations, which describe the precession of the electronic magnets. It will be recalled that in Part I, we looked at nuclear magnetic resonance sometimes as the turning-over of nuclear magnets and sometimes as an outcome of precession. Now we are going to treat the electronic resonance as an outcome of precession.

The fundamental vector equation, which was given in a sort of diluted form as equation (6) of Part I, reads as follows:

$$dp/dt = \mu_e \times H_i \quad (15)$$

Here p and μ_e stand for the angular momentum and the magnetic moment of the electron, and H_i for the field which operates on the electron. We have seen that μ_e/p is written as $ge/2mc$; we denote this quantity by γ ; and we give it the minus sign because, for the electron, angular momentum and magnetic moment are antiparallel to one another. Now we have:

$$d\mu_e/dt = -\gamma\mu_e \times H_i \quad (16)$$

This we proceed to write as three scalar equations; but first we replace μ_e by M . This will help to do away with the implication that the magnetic moment varies in magnitude (it is the *direction* that changes with

time) and will also convey the plausible suggestion that all of the resonating electrons in the substance are coupled parallel, so that M can signify the magnetization of the substance. We have:

$$\begin{aligned} dM_x/dt &= -\gamma(M_y H_{iz} - M_z H_{iy}) \\ dM_y/dt &= -\gamma(M_z H_{ix} - M_x H_{iz}) \\ dM_z/dt &= -\gamma(M_x H_{iy} - M_y H_{ix}) \end{aligned} \quad (17)$$

Now we are to make the following important substitutions, some of which are approximations.

(1) Presuming that M the magnetization of the substance will not deviate far from the z -direction, we are to write M for M_z .

(2) For H_{iz} , the z -component of the field actually operating upon the electrons, we are to write $(H - N_z M)$. Here H stands as heretofore for the applied field and N_z for the "demagnetizing factor" in the z -direction, which latter is a measure of the strength of the free poles on those surfaces of the sample which face the pole-pieces of the magnet (Fig. 1). Thus $-N_z M$ is the value of the field produced in the substance by these free poles.

(3) For H_{ix} and H_{iy} we are to write $-N_x M_x$ and $-N_y M_y$. This means that whatever applied fields there may be in the x and the y -directions are negligible, and yet the components of magnetization in these directions are not negligible, so that the free poles on the surfaces perpendicular to x and to y respectively are producing the internal fields of which $-N_x M_x$ and $-N_y M_y$ are the strengths.

(4) We are to ignore terms in which the product $M_x M_y$ appears, these being small.

The fourth of these conditions makes dM_z/dt vanish: we are left with only two of the three equations (17), a convenience. Making the substitutions allowed by the first three conditions, we find that the other two assume the forms:

$$\begin{aligned} dM_x/dt &= -\gamma M_y [H - (N_z - N_y)M] \\ dM_y/dt &= -\gamma M_x [-H + (N_z - N_x)M] \end{aligned} \quad (18)$$

Now suppose that M_x and M_y are periodic functions of time, of frequency ν . We write them as $M_x^0 \exp(2\pi i \nu t)$ and $M_y^0 \exp(2\pi i \nu t)$. Substituting into (18), we find:

$$\begin{aligned} 2\pi i \nu M_x^0 + \gamma [H + (N_y - N_z)M] M_y^0 &= 0 \\ -\gamma [H - (N_z - N_x)M] M_x^0 + 2\pi i \nu M_y^0 &= 0 \end{aligned} \quad (19)$$

These two simultaneous equations will be compatible with one another — one might say that they make sense — only if they are ultimately the same equation. The ratio of the coefficients of M_x^0 and M_y^0 in the one must be the same as the ratio of the corresponding coefficients in the other. In words more natural to algebraists, the determinant of the coefficients must vanish. It turns out that this condition determines a specific value of ν , and this value is the resonance-frequency:

$$\nu = (ge/4\pi mc) \sqrt{[H + (N_x - N_z)M][H + (N_y - N_z)M]} \quad (20)$$

For reasons deriving from the history of celestial mechanics, this procedure is known as “solving the secular equation.”

In the most common experimental set-up, the sample is a thin layer parallel to the z -direction — so thin that by comparison with its breadth, the free poles at the surfaces opposite the pole-pieces of the magnet may be regarded as infinitely far away. Under these conditions N_z vanishes, and so does N_x if we lay the x -axis parallel to the surface of the thin layer; but N_y does *not* vanish, it is in fact equal to 4π . Under the radical, the first factor becomes equal to H and the second to $H + 4\pi M$, which latter is by definition the induction B . We have:

$$\nu = (ge/4\pi mc) \sqrt{HB} \quad (21)$$

Note here that since B depends upon both H and M , one cannot use the formula unless one knows the value of M , which is the magnetization of the substance at saturation. This usually requires knowledge obtained from other experiments; but we shall meet with a case in which, at least “in principle,” the value of B may be found from the resonance-experiment itself.

Equation (21) is the commonest formula for the ferromagnetic resonance, for it fits the “geometry” of the original and of most of the subsequent experiments. Yet there are other formulae corresponding to other geometries, and two of these are particularly important.

It is feasible to orient the layer at right angles to the big applied field. For this case we shall do well to turn the axis of z so that it remains parallel to the big field. Now N_x and N_y vanish and N_z becomes 4π , and the formula is this:

$$\nu = (ge/4\pi mc)(H - 4\pi M) \quad (22)$$

The quantity $(H - 4\pi M)$ is the internal field H_i , the field strength within the magnetized body. This is the special case in which the right result is obtained by going back to equation (4) and putting for H the actual field strength at the scene of the resonating electrons. In other

words, this is the special case in which the naive approach does not lead the student astray.

A more singular special case is that of the sphere. In this case N_x and N_y and N_z are all three of them equal — equal to one another but not to zero. Nevertheless the formula is just our old formula (3), the same as though there were no magnetization at all:

$$\nu = (\mu/I)(H/h) = (ge/4\pi mc)H \quad (23)$$

One wonders how long it would have been before anyone set up equations (18) and derived equation (21), if all experiments had been performed with spheres.

In the foregoing pages we have derived the resonance frequency by making certain listed approximations in the basic equations (19). Among these approximations was the neglect of the oscillating field, parallel to the axis of x . We arrive at some interesting results by introducing this field into the equations and giving it an arbitrary frequency, while continuing to make all of the other approximations. It shall be denoted by $H_1 \exp(2\pi i \nu t)$; H_1 , it may be recalled, was the symbol used in Part I for the amplitude of this field. In this passage ν shall signify any frequency that the experimenter may choose to apply, while the resonance-frequency heretofore called ν shall change its symbol and become ν_0 .

On the right-hand side of the second of the equations (19) will now appear, as the reader can show for himself, $-\gamma M H_1$ instead of zero. The two simultaneous equations now make sense for any value of ν , instead of just the value ν_0 . On solving them for M_x^0 , one finds:

$$M_x^0/H_1 = \frac{M}{H + (N_x - N_z)M} \frac{1}{1 - (\nu/\nu_0)^2} \quad (24)$$

The quantity on the left, and hence also the quantity to which it is equated, is the "susceptibility" of the substance with respect to this oscillating field which, be it remembered, is imposed at right angles to the big applied field.

The quantity on the right has the well-known form of an optical dispersion-curve. Suppose the frequency to be increased from zero. The susceptibility rises from a finite and non-zero value at $\nu = 0$ to positive infinity at the resonance-frequency ν_0 ; here it jumps suddenly to negative infinity, from which value it rises asymptotically to zero as the frequency is increased toward infinity.

In magnetics there are methods of measuring directly, not the susceptibility χ itself but the sum $(1 + 4\pi\chi)$, which is called the "permeability" and is denoted by μ . It is evident that while the susceptibility

is rising, with increase of frequency, from its negative-infinite value at ν_0 to its asymptotic value of zero at infinite frequency, the permeability is rising from negative infinity to an asymptotic value which is equal to $+1$. Somewhere along this range of frequencies it must pass through zero, at a frequency to be denoted by ν_i . For a stratum parallel to both the big applied field and the oscillating field, it is easily shown that ν_i is equal to $(ge/4\pi mc)B$. This offers a way of determining B and consequently M .

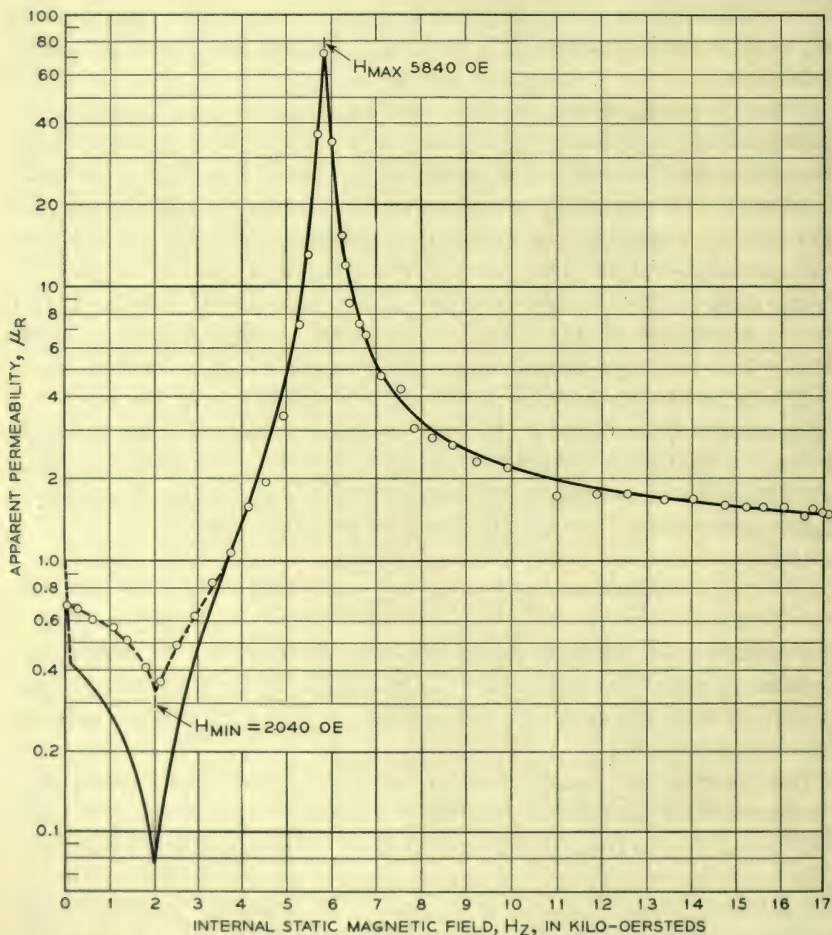


Fig. 4.—Ferromagnetic resonance in Heusler alloy (Cu-Mn-Al). “Apparent” permeability is plotted against H at constant frequency; the resonance-maximum (at the position corresponding to $g = 2.02$) is vividly shown, as is the minimum mentioned in the text. The solid curve is a theoretical curve based on a specific assumption about damping. (W. A. Yager and F. R. Merritt).

Next suppose that what is plotted against ν is not μ but $|\mu|$, the absolute value of the permeability. The portions of the μ -vs- ν curve which were below the horizontal axis now appear inverted and above the horizontal axis. The curve has an upward-pointing peak reaching to infinity at ν_0 , and a downward-pointing peak touching the axis with its tip at ν_i .

Such is the general aspect of the curve of Fig. 4, pertaining to a Heusler alloy. There are superficial differences: the curve of Fig. 4 is plotted against H for constant frequency, and the scale along the axis of ordinates is logarithmic. The reader can easily make allowance for these. There is also a fundamental difference: the curve reveals the presence of damping or relaxation, which broadens the peaks and prevents $|\mu|$ from rising to infinity or dropping quite to zero. The continuous curve is derived from a theory which involves a specific assumption about the damping; one sees that it agrees well with the data excepting in a region around the minimum. Curves such as these are likely to be influenced by anisotropy in the ferromagnetic substance, which reversely can be evaluated from the curves.

How about the values of g for ferromagnetic substances? The Heusler alloy to which Fig. 4 pertains has a value of g which, so far as the accuracy of the experiment permits us to judge, may be identical with the ideal value (the most probable value is however 2.01). This is an exception and not the rule. The range of values is rather wide, though apparently not so wide as in the strongly paramagnetic salts. Most of them lie between 2.22 (for cobalt) and 2.01 (for the Heusler alloy aforesaid); but there are instances of values still higher, including one of 3.75 for manganese arsenide. There is also at least one value lower than 2.00; it is presented by gadolinium, a very interesting element. Below its Curie point at 16° absolute, gadolinium shows a resonance-peak of which the breadth interferes with a precise location of its top; the value of g is given as 1.95 to 1.96. Above the Curie point, gadolinium is paramagnetic, but the peak persists and is sharper; the value of g is 1.95 ± 0.03 . I remind the reader that when the experiment is such that formula (21) must be used, a g -value implies an assumption about the value of M the magnetization of the substance at saturation.

I must not close this topic without alluding to something which there is not space to expound. Experiments on the "gyromagnetic effect" — something which has a much longer history than ferromagnetic resonance — lead to values of a quantity which has also been denoted by g . Until a few years ago it was supposed that this quantity must be the same as the g of these pages; but experiment has ruled otherwise, and theory has been successful in at least suggesting a reason. The g of these

pages is now called the spectroscopic splitting factor; the other has been set apart as the "gyromagnetic" g , and some people have even taken to writing it as g' , which seems rather unfair to the senior g . It seems to be a general rule that when one of the two is greater than 2.00 the other is smaller than 2.00; and in the case of the Heusler alloy, they may well coincide.

ACKNOWLEDGEMENT

Among the many who have helped me with this article I wish to extend especial thanks to Charles Kittel, A. F. Kip, W. D. Knight, M. E. Packard, E. M. Purcell, J. H. Van Vleck and W. A. Yager; and to Messrs. Packard, Purcell and B. Bleaney for prints of the illustrations captioned with their names.

REFERENCES

This makes no pretense of being a bibliography of magnetic resonance: such an enterprise would cover more pages of this Journal than these articles themselves. It is somewhat, but not much, more than a listing of the sources of the data quoted in the articles, with the names which have been omitted in the belief that names tend to slow down exposition. Many relevant papers, with an abundance of footnote references to anterior work, are to be found in what I abbreviate by P.I.C.S.R. ("Proceedings of the International Conference on Spectroscopy at Radiofrequencies, Amsterdam, 1950" separately published and also printed as part of Volume 17 of *Physica*).

The *locus classicus* for the nuclear resonance absorption is the paper of Bloembergen, Purcell and Pound (*Phys. Rev.* **73**, p. 670, 1948); the first publication of this school is by Purcell, Torrey and Pound in *Phys. Rev.* **69**, p. 37, 1946. The *locus classicus* for the precession-theory and the nuclear-induction method is the paper of Bloch (*Phys. Rev.* **70**, p. 460, 1946), followed by the first lengthy description of nuclear-induction measurements by Bloch, Hansen and Packard (*ibid.* p. 464); the first publication of this school is in *Phys. Rev.* **69**, p. 127, 1946. The discovery of nuclear magnetic resonance by the molecular-beam technique was disclosed by Rabi, Zacharias, Millman and Kusch in *Phys. Rev.* **53**, p. 318, 1938, and a more detailed account is given *ibid.* **55**, p. 526, 1939; consult also the paper of Kellogg, Ramsey, Rabi and Zacharias *ibid.* **57**, p. 677 (1940) for the resonances of protons and deuterons in molecular beams of H_2 , D_2 and HD .

The reference to the article of G. E. Pake (*Am. Jour. Phys.* **18**, pp. 438-452 and pp. 473-486, 1950) is here repeated to draw attention to this excellent survey of nuclear magnetic resonance and relaxation. Another survey article is that of Rollin, *Reports on Recent Progress in Physics*, 1948-49. The reference for the chemical shift in ethyl alcohol (Fig. 7 of Part I) is Arnold, Dharmatti and Packard, *J. Chem. Phys.* **19**, p. 507, 1951. For the influence of F-centres on nuclear relaxation-time see Hatton and Rollin, *Proc. Roy. Soc.* **198**, p. 231, 1949.

For the final experiments on determination of g for electrons in hydrogen atoms see Koenig, Prodell and Kusch (*Phys. Rev.* **88**, p. 191, 1952); references to earlier

work on other atoms are listed there. For the determination of the proton moment see Gardner, *Phys. Rev.* **83**, p. 996, 1951, and Sommers, Hipple and Thomas, *ibid.* **80**, p. 487, 1950. For the measurement of g with F-centres see Hutchinson and Noble, *ibid.* **87**, p. 1152, 1952, and Tinkham and Kip. *ibid.* **83**, p. 657, 1951, and Schneider and England in P.I.C.S.R.; this last is also the source for the work on zinc sulfide phosphor. For resonance of conduction-electrons see Griswold, Kip and Kittel, *Phys. Rev.* **88**, p. 951, 1952, and papers yet to be published.

The resonance-peak of Fig. 2 of Part II of this article, for porphyraxide, comes from Holden, Yager and Merritt, *J. Chem. Phys.* **19**, p. 1319, 1951. The literature of resonance in strongly paramagnetic salts is extensive and tough; I refer to the papers in P.I.C.S.R. and the references they give. The basic theory is to be found in the book of J. H. Van Vleck, *Electric and Magnetic Susceptibilities* (Oxford, 1932).

Hyperfine structure of electron resonance was discovered by the late R. P. Penrose (see *Nature*, **163**, pp. 988 and 992, 1949). This field is almost a monopoly of Britain and in particular of Oxford; many of the papers bear the name of Bleaney with or without collaborators. Fig. 3 of Part II of this article comes from *Proc. Phys. Soc.* **63**, p. 1369, 1950; the statements about vanadium 50 from *ibid.* **65**, p. 952, 1952.

The discoverer of ferromagnetic resonance was J. H. E. Griffiths (see *Nature*, **158**, p. 670, 1946). The precession-theory for ferromagnetic substances is due to Kittel; equation (21) of this article is derived in *Phys. Rev.* **71**, p. 270, 1947; a fuller treatment appears *ibid.* **73**, p. 155, 1948. Survey articles are those of Van Vleck in P.I.C.S.R. and Kittel in *Jour. de Phys.*, **12**, p. 291, 1951. Fig. 4 comes from the paper of Yager and Merritt, in *Phys. Rev.* **73**, p. 318, 1949; a similar curve for supermalloy appears in the preceding paper; references to the g -values of other ferromagnetic substances are given by Yager. The question of " g versus g " is discussed by Kittel in *Phys. Rev.* **76**, p. 743 (1949).

A Study of Non-Blocking Switching Networks

By CHARLES CLOS

(Manuscript received October 30, 1952)

This paper describes a method of designing arrays of crosspoints for use in telephone switching systems in which it will always be possible to establish a connection from an idle inlet to an idle outlet regardless of the number of calls served by the system.

INTRODUCTION

The impact of recent discoveries and developments in the electronic art is being felt in the telephone switching field. This is evidenced by the fact that many laboratories here and abroad have research and development programs for arriving at economic electronic switching systems. In some of these systems, such as the ECASS System,* the role of the switching crossnet array becomes much more important than in present day commercial telephone systems. In that system the common control equipment is less expensive, whereas the crosspoints which assume some of the control functions are more expensive. The requirements for such a system are that the crosspoints be kept at a minimum and yet be able to permit the establishment of as many simultaneous connections through the system as possible. These are opposing requirements and an economical system must of necessity accept a compromise. In the search for this compromise, a convenient starting point is to study the design of crossnet arrays where it is always possible to establish a connection from an idle inlet to an idle outlet regardless of the amount of traffic on the system. Because a simple square array with N inputs, N outputs and N^2 crosspoints meets this requirement, it can be taken as an upper design limit. Hence, this paper considers non-blocking arrays where less than N^2 crosspoints are required. Specifically, this paper describes for an implicit set of conditions, crossnet arrays of three, five,

* Malthaner, W. A., and H. Earle Vaughan, An Experimental Electronically Controlled Switching System. Bell Sys. Tech. J., 31, pp. 443-468, May, 1952.

etc., switching stages where less than N^2 crosspoints are required. It then deals with conditions for obtaining a minimum number of crosspoints, cases where the N inputs and N outputs can not be uniformly assigned to the switches, switching arrays where the inputs do not equal the outputs, and arrays where some or all of the inputs are also outputs.

SQUARE ARRAY

A simple square array having N inputs and N outputs is shown in Fig. 1. The number of crosspoints equals N^2 and any combination of N or less simultaneous connections can exist without blocking between the inputs and the outputs. The number of switching stages, s , is equal to 1. The number of crosspoints, $C(s)$, is:

$$C(1) = N^2 \quad (1)$$

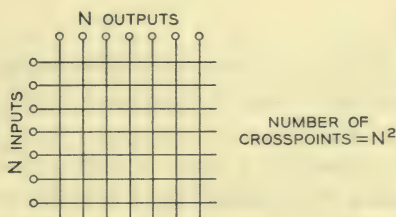


Fig. 1 — Square Array.

THREE-STAGE SWITCHING ARRAY

An array where less than N^2 crosspoints are required is shown in Fig. 2. This array has $N = 36$ inputs and $N = 36$ outputs. There are three switching stages, namely, an input stage (a), an intermediary stage (b), and an output stage (c). In stage (a) there are six 6×11 switches; in stage (b) there are eleven 6×6 switches; and in stage (c) there are six 6×11 switches. In total, there are 1188 crosspoints which are less than the 1296 crosspoints required by equation (1).

Of interest are the derivations of the various quantities and sizes of switches. In stage (a) the number, n , of inputs per switch was assumed to be equal to $N^{1/2}$, thus giving six switches and six inputs per switch. In a similar manner stage (c) was assigned six switches and six outputs per switch. The number of switches required in stage (b) must be sufficient to avoid blocking under the worst set of conditions. The worst case occurs when between a given switch in stage (a) and a given switch in stage (c): (1) five links from the switch in stage (a) to five correspond-

ing switches in stage (b) are busy; (2) five links from the switch in stage (c) are busy to five additional switches in stage (b); and (3) a connection is desired between the given switches. Thus eleven switches are required in stage (b). The remaining requirements, namely, eleven verticals per switch in stages (a) and (c) and six by six switches in stage (b) are then easily derived.

The number of crosspoints required for three stages, where $n = N^{1/2}$, is summarized by the following formula:

$$C(3) = (2N^{1/2} - 1)(3N) \quad (2)$$

$$= 6N^{3/2} - 3N \quad (2a)$$

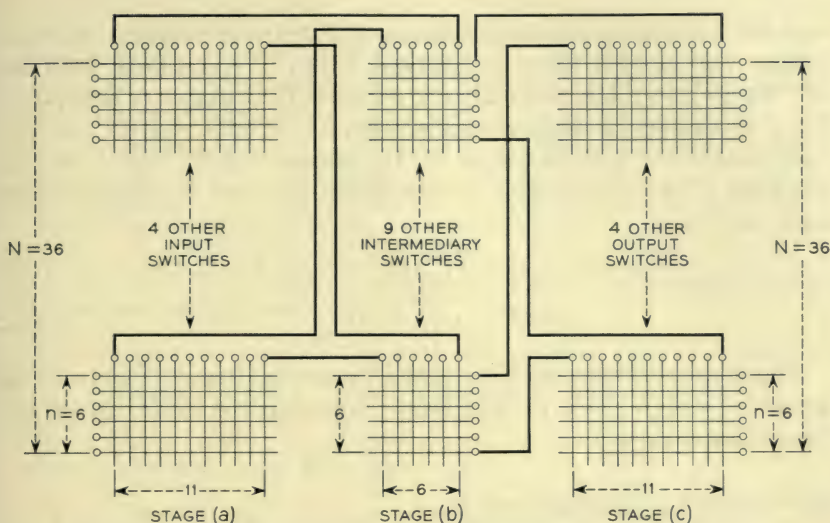
In Table I it may be noted that the number of crosspoints is less than N^2 for all cases of $N \geq 36$.

PRINCIPLE INVOLVED

The principle involved for determining the number of switches required in the intermediary stage is illustrated in Fig. 3. The figure is for a specific case from which one can generalize for n inputs on a given input switch and m outputs on a given output switch. In the figure it is desired to establish a connection from input B to output H . A sufficient number of intermediary switches are required to permit the $(n - 1)$ inputs other than B on the particular input switch and the $(m - 1)$ outputs other than H on the particular output switch to have connections to separate intermediary switches plus one more switch for the desired connection between B and H . Thus $n + m - 1$ intermediary switches are required.

TABLE I — CROSSPOINTS FOR SEVERAL VALUES OF N

N	Square Array N^2	Three-Stage Array $6N^{3/2} - 3N$
4	16	36
9	81	135
16	256	336
25	625	675
36	1,296	1,188
49	2,401	1,911
64	4,096	2,880
81	6,561	4,131
100	10,000	5,700
...	...	...
1,000	1,000,000	186,737
10,000	100,000,000	5,970,000



NUMBER OF CROSSPOINTS = $6N^{3/2} - 3N$ (1188 CROSSPOINTS WHEN $N=36$)

Fig. 2 — Three-stage switching array.

FIVE-STAGE SWITCHING ARRAY

A five-stage switching array is illustrated in Fig. 4. The analysis of this array can be made in the following manner. Each input and output switch is assumed to have $n = N^{1/3}$ inputs or outputs, respectively. Connection between a given input switch and a given output switch

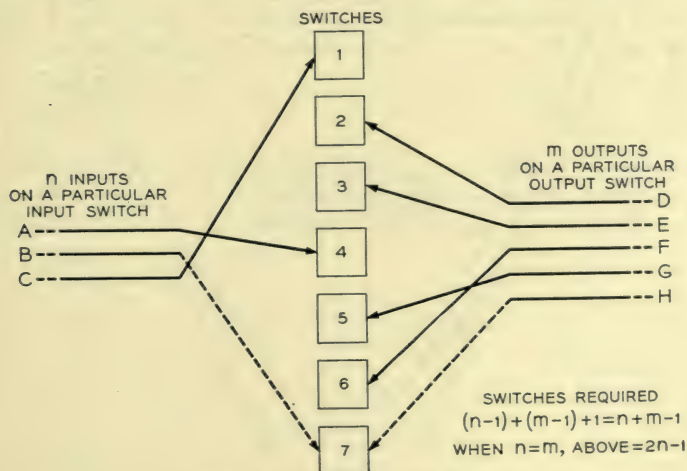


Fig. 3 — Principle involved.

is made via levels, a level consisting of three intermediary switching stages. The number of levels required is $(2N^{1/3} - 1)$. Each level has $N^{2/3}$ inputs and the same number of outputs. The number of crosspoints for a three-stage non-blocking array for $N^{2/3}$ inputs and $N^{2/3}$ outputs can be obtained from equation (2) by substituting $N^{2/3}$ for N in that equation. The total number of crosspoints required for the five-stage array is:

$$C(5) = (2N^{1/3} - 1)^2 3N^{2/3} + (2N^{1/3} - 1) 2N \quad (3)$$

$$= 16N^{4/3} - 14N + 3N^{2/3} \quad (3a)$$

The number of crosspoints required for several sizes of the five-stage array is given in Table II. The results are compared to the square and three-stage arrays.

SEVEN-STAGE SWITCHING ARRAY

A seven-stage switching array can be analyzed by considering paths requiring five intermediary switching stages as paths via switching aggregates. The number of such aggregates is $(2N^{1/4} - 1)$. Each aggregate has $N^{3/4}$ inputs and a like number of outputs. From equation (3) the crosspoints for each aggregate can be obtained by substituting

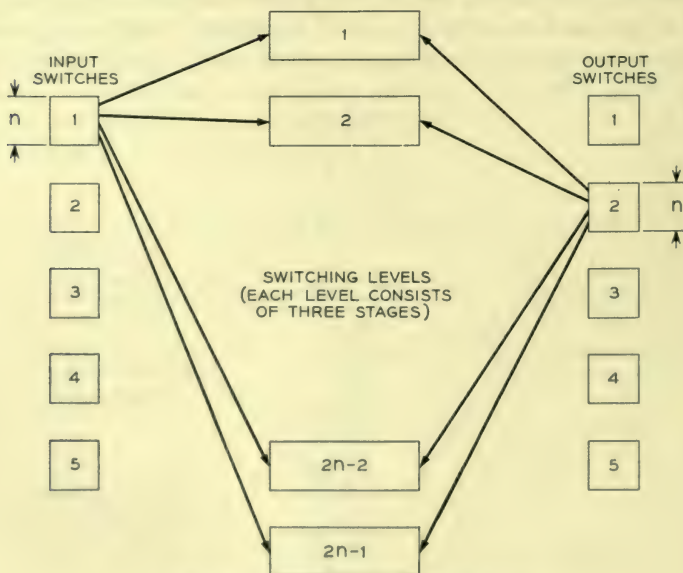


Fig. 4 — Five-stage switching array.

TABLE II — CROSSPOINTS FOR SEVERAL VALUES OF N

N	Square Array	Three-Stage Array	Five-Stage Array
64	4,096	2,880	3,248
729	531,441	115,911	95,013
1,000	1,000,000	186,737	146,300
10,000	100,000,000	5,970,000	3,434,488

$N^{3/4}$ for N in that equation. The total number of crosspoints required for the seven-stage array is:

$$C(7) = (2N^{1/4} - 1)^3 3N^{1/2} + (2N^{1/4} - 1)^2 2N^{3/4} + (2N^{1/4} - 1)2N \quad (4)$$

$$= 36N^{5/4} - 46N + 20N^{3/4} - 3N^{1/2} \quad (4a)$$

GENERAL MULTI-STAGE SWITCHING ARRAY

Equations (1), (2a), (3a) and (4a) are herewith tabulated as a series of polynomials together with the next polynomial:

$$C(1) = N^2 \quad (1)$$

$$C(3) = 6N^{3/2} - 3N \quad (2a)$$

$$C(5) = 16N^{4/3} - 14N + 3N^{2/3} \quad (3a)$$

$$C(7) = 36N^{5/4} - 46N + 20N^{3/4} - 3N^{1/2} \quad (4a)$$

$$C(9) = 76N^{6/5} - 130N + 86N^{4/5} - 26N^{3/5} + 3N^{2/5} \quad (5)$$

These polynomials can be determined for any number of switching stages from the following formula where s is an odd integer:

$$C(s) = 2 \sum_{k=\frac{s+1}{2}}^{\frac{s+1}{2}} N^{\frac{2k}{s+1}} \left(2N^{\frac{2}{s+1}} - 1 \right)^{\frac{s+3}{2}-k} + N^{\frac{4}{s+1}} \left(2N^{\frac{2}{s+1}} - 1 \right)^{\frac{s-1}{2}} \quad (6)$$

An alternative expression equivalent to equation (6) has been suggested by S. O. Rice and J. Riordan. The recurrence relation used in individually deriving the foregoing polynomials can be used to directly derive the following formula:

$$C(2t+1) = \frac{n^2(2n-1)}{n-1} [(5n-3)(2n-1)^{t-1} - 2n^t] \quad (6a)$$

where $s = 2t + 1$

$$N = n^{t+1}$$

Table III gives comparative numbers of crosspoints for various num-

TABLE III — CROSSPOINTS FOR VARIOUS NUMBERS OF SWITCHING STAGES, s , AND VALUES OF N

N	$s = 1$	$s = 3$	$s = 5$	$s = 7$	$s = 9$
100	10,000	5,700	6,092	7,386	9,121
200	40,000	16,370	16,017	18,898	23,219
500	250,000	65,582	56,685	64,165	78,058
1,000	1,000,000	186,737	146,300	159,904	192,571
2,000	4,000,000	530,656	375,651	395,340	470,292
5,000	25,000,000	2,106,320	1,298,858	1,295,294	1,511,331
10,000	100,000,000	5,970,000	3,308,487	3,159,700	3,625,165

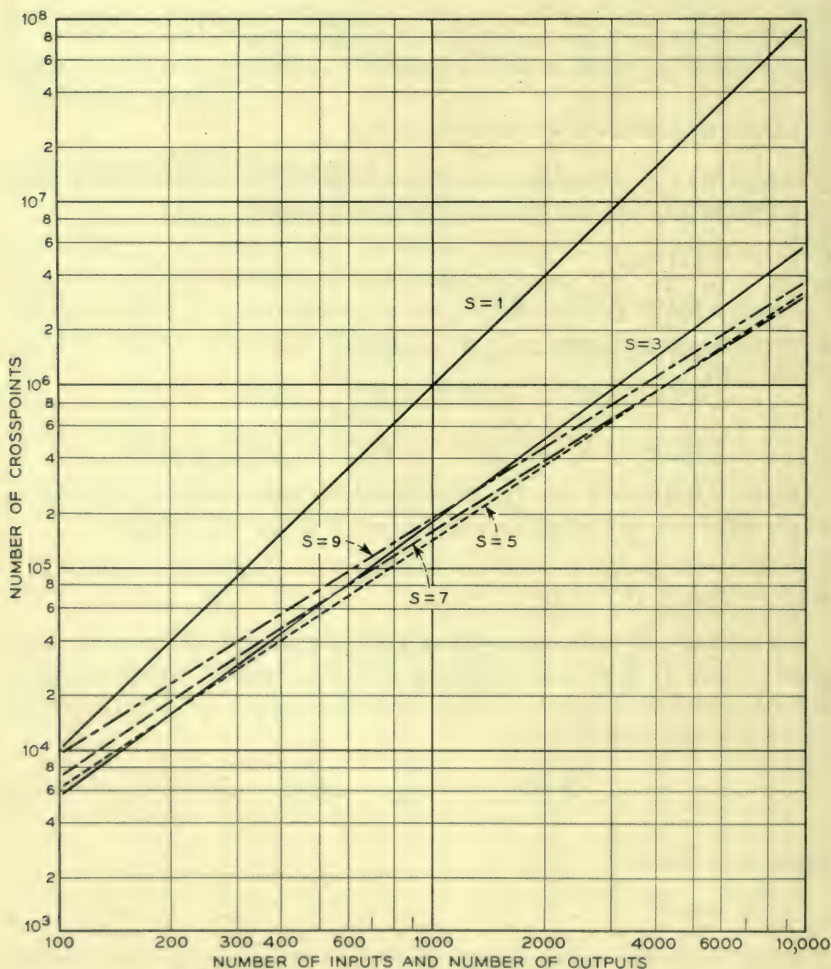


Fig. 5 — Crosspoints versus switching stages.

bers of switching stages and sizes of N . The data of Table III are plotted on Figure 5. The series of curves appear to be bounded by an envelope, representing a minimum of crosspoints. The next section dealing with minima indicates that points exist below this envelope.

MOST FAVORABLE SIZE OF INPUT AND OUTPUT SWITCHES IN THE THREE-STAGE ARRAY

The foregoing derivations were for implicit relationships between n and N , namely, n being the $\left(\frac{s+1}{2}\right)$ th root of N . To obtain minimum number of crosspoints a more general relationship is required. For the three stage switching array this is:

$$C(3) = (2n - 1) \left(2N + \frac{N^2}{n^2} \right) \quad (7)$$

When $n = N^{1/2}$ equation (7) reduces to equation (2).

For a given value of N , the minimum number of crosspoints occurs when $dC/dn = 0$ which gives:

$$2n^3 - nN + N = 0 \quad (8)$$

This equation has the following two pairs of integral values:

$$n = 2, \quad N = 16 \quad \text{and} \quad n = 3, \quad N = 27$$

As N approaches large values equation (8) can be approximated by:

$$N \doteq 2n^2 \quad (9)$$

Graphs of equations (8) and (9) are shown in Fig. 6. In Table IV the numbers of crosspoints are based on the nearest integral values of n for given values of N .

Where comparisons can be made, Table IV indicates fewer crosspoints than does Table I. This fact can be realized in another manner. By eliminating n in equations (7) and (9), the result for large values of N is:

$$C(3) \doteq 4 (2)^{1/2} N^{3/2} - 4N \quad (10)$$

Equation (10) indicates fewer crosspoints than does equation (2).

MOST FAVORABLE SWITCH SIZES IN THE FIVE-STAGE ARRAY

If n be the number of inputs per input switch and outputs per output switch, and m be the number of inputs per switch in the second stage

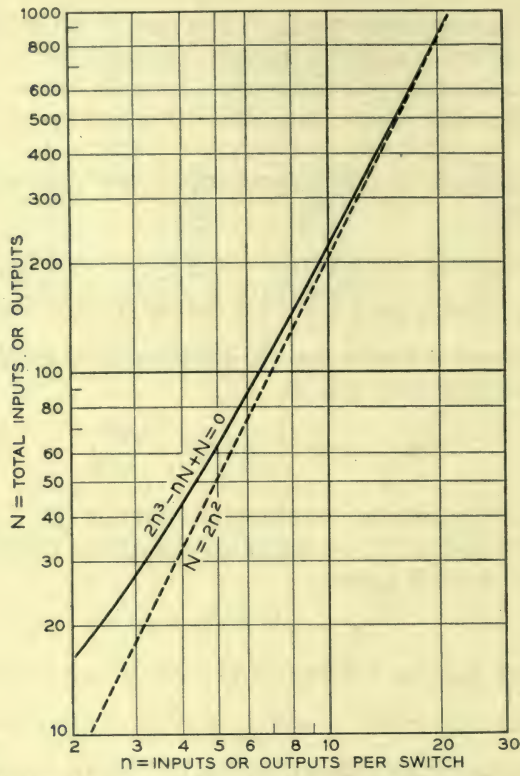


Fig. 6 — Relationship between N and n for minima in crosspoints. Three-stage array.

TABLE IV — CROSSPOINTS FOR SEVERAL VALUES OF N

N	Nearest Integral Value of n	Number of Crosspoints	
		N^2	Equation (7)
16	2	256	288
27	3	729	675
40	4	1,600	1,260
44	4	1,936	1,463
55	5	3,025	2,079
60	5	3,600	2,376
65	5	4,225	2,691
78	6	6,084	3,575
84	6	7,056	4,004
98	7	9,604	5,096
105	7	11,025	5,655

and outputs per switch in the fourth stage, then the following equation gives the total number of crosspoints:

$$C(5) = (2n - 1) \left[2N + (2m - 1) \left(\frac{2N}{n} + \frac{N^2}{n^2 m^2} \right) \right] \quad (11)$$

The partial derivative of this equation with respect to m when set equal to zero yields:

$$n = \frac{N(m - 1)}{2m^3} \quad (12)$$

The partial derivative of this equation with respect to n when set equal to zero yields the following equation:

$$N = \frac{nm^2(2n^2 + 2m - 1)}{(2m - 1)(n - 1)} \quad (13)$$

Equations (12) and (13) can be solved for n and m in terms of given values of N . For example for $N = 240$, we obtain $n = 6.81$ and $m = 3.56$.

SEARCH FOR THE SMALLEST N FOR A GIVEN n FOR THE THREE-STAGE ARRAY

For a given value of n , equation (7) furnishes a means for locating that size of three-stage switching array which has N^2 or fewer crosspoints. This can be done by setting equation (7) equal to N^2 :

$$N^2 = (2n - 1) \left(2N + \frac{N^2}{n^2} \right) \quad (14)$$

and solving for N in terms of n . The solution is:

$$N \geq \frac{2n^2(2n - 1)}{(n - 1)^2} \quad (15)$$

Minimum values of N for given values of n are listed in Table V. This table also lists the next highest N exactly divisible by n . From this table it appears that when $N = 24$, we have the smallest switching array for which it may be possible to have less than N^2 crosspoints. However for $N = 25$, as shown in Table I, equation (2) gives more than N^2 crosspoints. The problem is one of finding an array for $N = 25$ with fewer than N^2 crosspoints. For this and all cases beyond, the next section indicates that it is profitable to consider situations where N is not exactly divisible by n .

CASES IN THE THREE-STAGE SWITCHING ARRAY WHERE $N \equiv r(\text{MOD } n)$

Table I indicated that for $N = 25$ and $n = 5$ a total of 675 crosspoints were required. A square array requires only 625. Fig. 7 shows a layout of switches where $N = 25$ and $n = 3$. In this case one input is left over when 25 inputs are divided into threes. The lone input requires three paths to the intermediary switches. This is in accordance with Fig. 3. The lone output also requires three paths to the intermediary switches. Also from Fig. 3, the lone input to the lone output requires only one path. Hence there must be one switch capable of connecting the lone input to the lone output. The number of crosspoints required is 615 which is less than the 625 required by the square array. This scheme can be extended to any case where $N = kn + r$, where the remainder, r , is an integer greater than zero but less than n . The formula for the number of crosspoints where k input and k output switches of size n and one input and output switch of size r are used is:

$$C = 2(2n - 1)(N - r) + 2(n + r - 1)r + (n - r) \left(\frac{N - r}{n} \right)^2 + (n + r - 1) \left(\frac{N - r}{n} + 1 \right)^2 - n + r \quad (16)$$

I. G. Wilson has pointed out that for a lone input the crosspoints in the intermediary switches can be used to isolate its possible connections hence no crosspoints are required in the input stage. This likewise applies for a lone output. With this modification the array in Fig. 7 requires six fewer crosspoints. For this case, when $r = 1$, the number of crosspoints is:

$$C = 2(2n - 1)(N - 1) + (n - 1) \left(\frac{N - 1}{n} \right)^2 + n \left(\frac{N - 1}{n} + 1 \right)^2 - n + 1 \quad (16a)$$

TABLE V — MINIMUM VALUES OF N FOR GIVEN VALUES OF n

n	N per Equation 15	$N \equiv 0 \pmod{n}$
2	24	24
3	22.5	24
4	24.9	28
5	28.1	30
6	31.7	36

J. Riordan has found a more efficient arrangement for cases where $N = kn + r$. In place of using k switches of size n and one switch of size r , he proposes that $(k + 1 - n + r)$ switches of size n and $(n - r)$ switches of size $n - 1$ be used. For this case the number of crosspoints is:

$$C = 2(2n - 1)(k + 1 - n + r)n + 2(2n - 2)(n - r)(n - 1) + (2n - 3)(k + 1)^2 + 2(k + 1)(k + 1 - n + r) \quad (17)$$

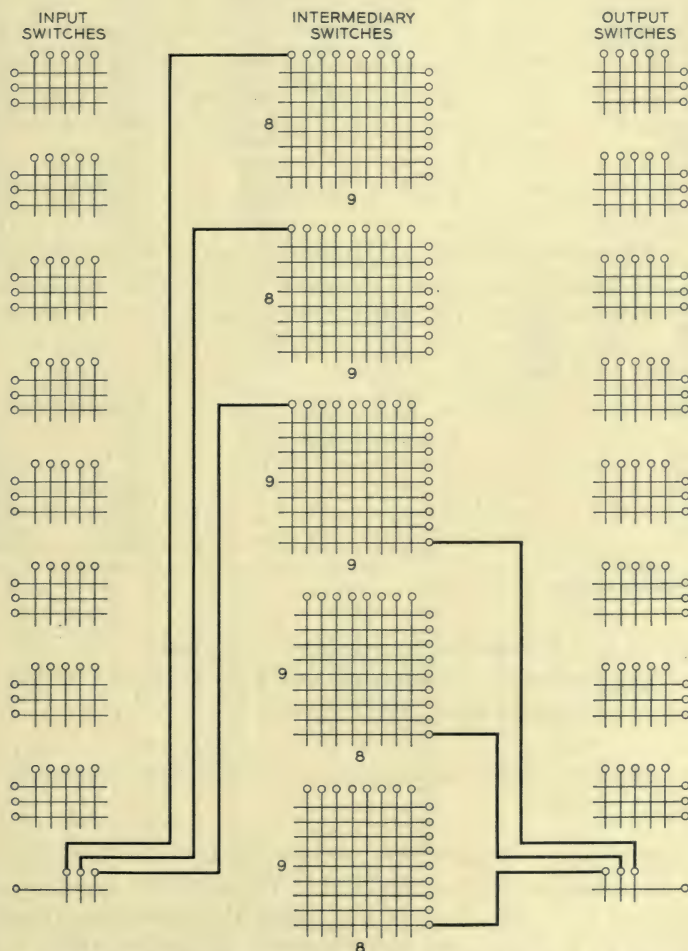


Fig. 7 — Three-stage array. $25 \equiv 1 \pmod{3}$. An equivalent arrangement is to provide two 8×8 and three 9×9 intermediary switches. Two of the 9×9 switches need only 80 crosspoints.

TABLE VI — CROSSPOINTS FOR VARIOUS VALUES OF N AND n

N	Square Array	Three-Stage Array			
		$n = 2$	$n = 3$	$n = 4$	$n = 5$
23	529*	540	530	556	—
24	576	576	560*	588	—
25	625	625	609*	633	—
26	—	663	643*	667	—
27	—	716	675*	701	—
28	—	756	730*	735	—
29	—	813	766*	788	—
30	—	855	800*	824	864
31	—	—	861	860*	911
32	—	—	899	896*	951
33	—	—	935*	957	991
34	—	—	1002	995*	1031
35	—	—	1042	1033*	1071
36	—	—	1080	1071*	1128
37	—	—	1153	1140*	1170
38	—	—	1195	1180*	1212
39	—	—	1235	1220*	1254
40	—	—	1314	1260*	1296
		$n = 4$	$n = 5$	$n = 6$	$n = 7$
50	1819	1800*	1879	—	—
60	2415	2376*	2420	—	—
70	3164	3024*	3056	—	—
80	3920	3744*	3764	—	—
90	—	4536	4455*	4499	—
100	—	5400	5291*	5315	—
110	—	—	6199	6100	6156
120	—	—	7040	7044	6975*
130	—	—	8076	7923*	7947
140	—	—	—	8840*	8860
150	—	—	—	9968	9811*
160	—	—	—	10979	10800*

* Minimum values.

Equation (17) is identical to equation (16) when $r = n - 1$. There are two cases, namely, when $n = 2$ and $n = 3$ where equation (16a) gives fewer crosspoints than does equation (17).

SEARCH FOR THE MINIMUM NUMBER OF CROSSPOINTS BETWEEN $N = 23$ AND $N = 160$

The equations of the preceding sections furnish a means for searching for minimum crossnet arrays. Table VI shows the results of such a search up to $N = 160$. Results are indicated in unit steps from $N = 23$ to $N = 40$ and for every tenth interval thereafter. At $N = 161$, a five-stage array requires the fewest crosspoints.

Table VI was computed by the use of finite differences. The equations

were:

$$C[(k+1)n] - C(kn) = (2n-1)(2n+2k+1) \quad (18)$$

$$C(kn+r+1) - C(kn+r) = 2(k+3n-1) \quad (19)$$

$$C(kn+1) - C(kn) = 2kn+1 \quad (19a)$$

Equation (18) was derived from equation (7) with N being replaced by $(k+1)n$ and by kn as required. Equation (19) was derived from equation (17) with r being replaced by $r+1$ as required. This equation applies for all values of n greater than 3 and for the particular case of $n=3$ and $r=2$. Equation (19a) was derived from Equations (16a) and (7) and is for the particular case of $r=1$, when $n=2$ and $n=3$.

SEARCH FOR THE MINIMUM NUMBER OF CROSSPOINTS FOR $N = 240$

For a case where N is large enough to require five-switching stages, the search for the minimum number of crosspoints should be based on equations (12) and (13) and on the use of Table VI. The method is suggested by means of Table VII. The data in a previous section indicate

TABLE VII — CROSSPOINTS FOR $N = 240$ AND VARIOUS VALUES OF n

Input and Output Stages				Intermediary Stages				Total Crosspoints
n	No. of Switches	Size of Switches	Cross-points	No. of Levels	Inputs and Outputs	m	Cross-points	
2	120	2 x 3	1,440	3	120 x 120*	8	20,925	22,365
3	80	3 x 5	2,400	5	80 x 80*	5	18,720	21,120
4	60	4 x 7	3,360	7	60 x 60*	5	16,632	19,992
5	48	5 x 9	4,320	9	48 x 48	4	15,120	19,440
6	40	6 x 11	5,280	11	40 x 40*	4	13,860	19,140
7	30	7 x 13	5,460	2	30 x 35	3	1,826	19,369
	5	6 x 12	720	11	35 x 35*	4	11,363	
8	30	8 x 15	7,200	15	30 x 30*	3	12,000	19,200
9	24	9 x 17	7,344	2	24 x 27	3	1,230	19,467
	3	8 x 16	768	15	27 x 27*	3	10,125	
10	24	10 x 19	9,120	19	24 x 24*	3	10,640	19,760
11	20	11 x 21	9,240	2	20 x 22	—	880	20,116
	2	10 x 20	800	19	22 x 22	—	9,196	
12	20	12 x 23	11,040	23	20 x 20	—	9,200	20,240
Crosspoints per equation (3) five-stage array.....								20,596
Crosspoints per equation (2) three-stage array.....								21,624
Crosspoints per equation (1) square array.....								57,600

* See Table VI for minimum number of crosspoints.

that a minimum should occur for $N = 240$, when $n = 6.81$ and $m = 3.56$. In Table VII the minimum occurs when $n = 6$ and $m = 4$. It fails to occur at $n = 7$ because 240 is not exactly divisible by 7. Except for this situation, the minimum would have occurred as predicted.

RECTANGULAR ARRAY

Referring to Fig. 1, if there were N inputs and M outputs, a simple rectangular array would result which would be capable of sustaining up to N or M , whichever is the lesser, simultaneous connections without blocking. The number of crosspoints is:

$$C(1) = NM \quad (20)$$

N INPUTS AND M OUTPUTS IN A THREE-STAGE ARRAY

For the case of a three-stage switching array with N inputs and M outputs, let there be n inputs per input switch and m outputs per output switch. A particular input to be able to connect without blocking under the worst set of conditions to a particular output will require $(n - 1) + (m - 1) + 1$ available paths. Thus by providing for that many intermediary switches, a non-blocking switching array is obtained. The number of crosspoints is:

$$C(3) = (n + m - 1) \left[N + M + \frac{NM}{nm} \right] \quad (21)$$

Differentiating this equation first with respect to n and then to m yields two partial differential equations whose solution indicates that a minimum is reached when $n = m$. Replacing m by n in equation (21), the equation for the number of crosspoints becomes:

$$C(3) = (2n - 1) \left[N + M + \frac{NM}{n^2} \right] \quad (22)$$

Solving for the minimum number of crosspoints gives the following expression:

$$n^3 - \frac{NM}{N + M} n + \frac{NM}{N + M} = 0 \quad (23)$$

When $N = M$ this equation reduces to equation (8).

The three-way relationships of n , N and M are shown in Fig. 8.

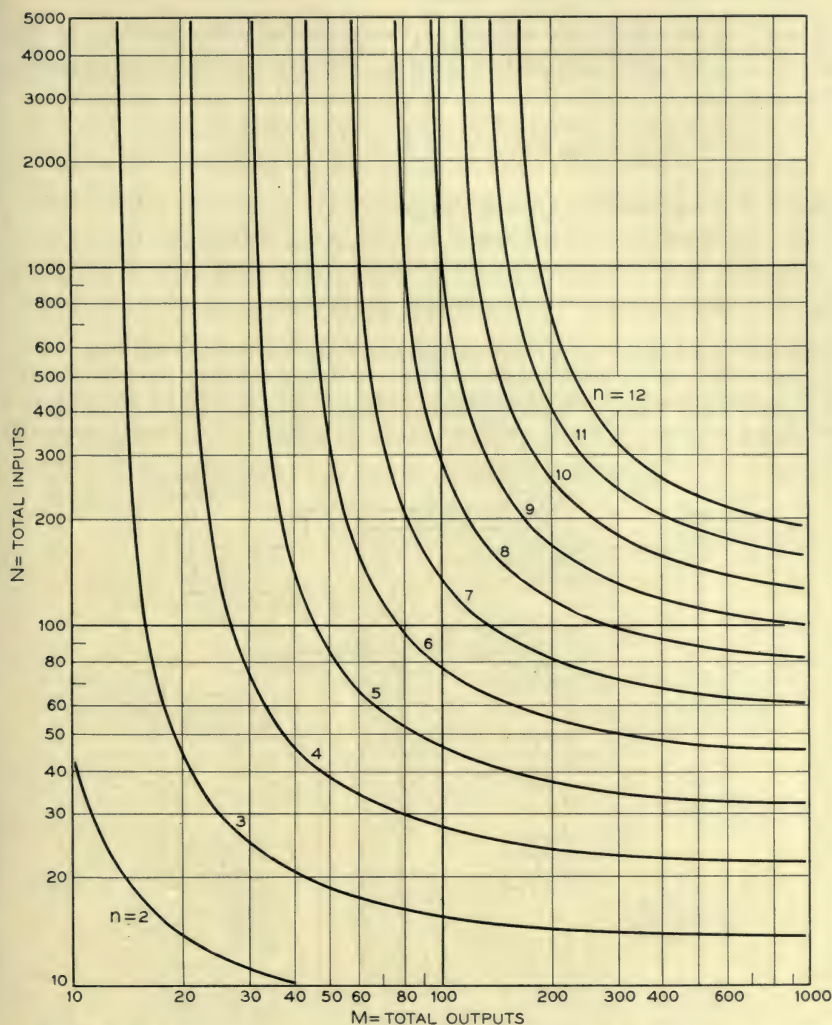


Fig. 8 — Relationship of n to N inputs and M outputs for a minimum in cross-points in a three stage array.

TRIANGULAR ARRAY

If a case exists where all inputs are also the outputs, then an arrangement such as is shown in Fig. 9 can be used. The crosspoints in the intermediary switches permit connections between all switches on the left hand side. For connections between two trunks on the same switch it is assumed that one of the links to an intermediary switch can

be used to establish the connection but without affecting any of the crosspoints on the intermediary switch. The number of crosspoints for this case is:

$$C = (2n - 1) \left(T + \frac{T^2}{2n^2} - \frac{T}{2n} \right) \quad (24)$$

where T = number of two-way trunks.

By differentiation, conditions for obtaining minimum numbers of crosspoints can be determined. The arrangement can also be extended to cases where extra switching stages are required.

ONE-WAY INCOMING, ONE-WAY OUTGOING AND TWO-WAY TRUNKS

A combination of the triangular array of Fig. 9 and of unequal inputs and outputs is shown in Fig. 10. In this figure, one-way incoming,

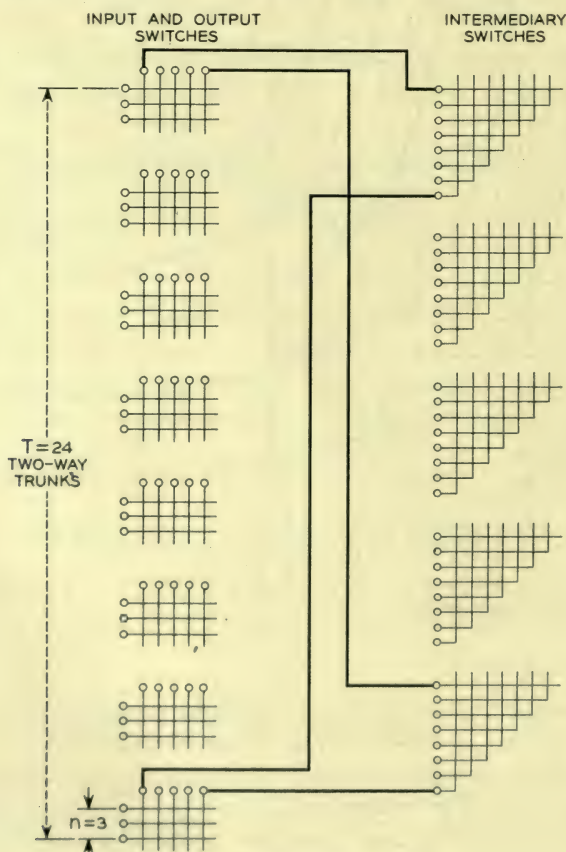


Fig. 9 — Triangular array.

one-way outgoing and two-way trunks can be freely interconnected without blocking. The number of crosspoints for this case is:

$$C = (2n - 1) \left[N + T + M + \frac{NT}{n^2} + \frac{MT}{n^2} + \frac{T^2}{2n^2} - \frac{T}{2n} \right] \quad (25)$$

The comments concerning the triangular array also apply for this case.

COMPARISON WITH EXISTING NETWORKS

Few existing crossnet arrays are non-blocking. An example is the four-wire intertoll trunk concentrating system. In one of its standard sizes 4,000 crosspoints are required for 100 incoming trunks and 40 outgoing intertoll trunks. From Fig. 8, for $N = 100$ and $M = 40$ it may be noted that the nearest integral value for n is 5. By substituting this value in equation (22), a non-blocking three-stage switching array of 2,700 cross-

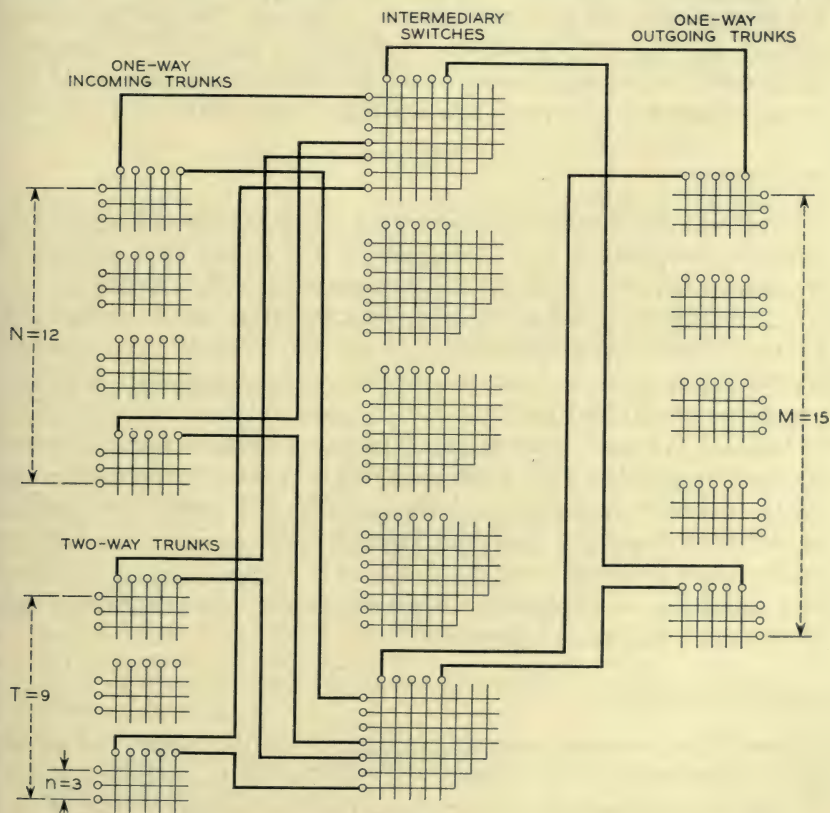


Fig. 10 — One-way incoming, one-way outgoing and two-way trunks.

points is found which could be used for the concentrating switch. In this case the new approach to the switching network problem may prove to be of value.

Comparisons with existing arrays having blocking are likely to be unfavorable because the grades of service are not the same. For instance, a No. 1 crossbar district-to-office layout of 1,000 district junctors and 1,000 trunks requires 80,000 crosspoints. This layout can handle 708 erlangs with a blocking loss of 0.0030. The minimum number of crosspoints with a non-blocking array is slightly less than 138,000. This, however, can handle 1,000 erlangs without blocking. By introducing blocking into the design methods described in this paper, a more favorable comparison with existing arrays having blocking can be made. This can be done by omitting certain of the paths. If done to an array requiring 1,000 inputs and 1,000 outputs a layout can be obtained requiring 79,900 crosspoints with a blocking loss of 0.0022 for a load of 708 erlangs. For this example, at least, it appears that the new design methods may prove to be valuable especially for use in the development of electronic switching systems where the control mechanism may not be dependent upon the particular switching array used.

CONCLUSION

In present day commercial telephone systems the use of non-blocking switching networks is rare. This may be due to the large number of crosspoints required. With the design methods described herein, a wider use of non-blocking networks may occur in future developments. For the usual case of networks with blocking, new systems have generally been designed by an indirect process. Several types and sizes of switching arrays are studied until the most economical one for a given level of blocking is found. With the new design methods, a straightforward approach is possible. Fig. 5 indicates that a region of minimum values exists. By first designing a non-blocking system with a reasonable number of switching stages and then omitting certain of the paths, the designer can arrive at a network with a given level of blocking and be very close to a minimum in crosspoints. The possibility of the adoption of this direct design method is important.

ACKNOWLEDGEMENT

In addition to those specifically mentioned in this paper, the author is also indebted to E. B. Ferrell, B. D. Holbrook, C. A. Lovell and E. F. Moore for suggestions and encouragement in the preparation of this paper.

The Evaluation of Wood Preservatives Part II

By REGINALD H. COLLEY

(Manuscript received September 22, 1952)

This paper offers a review and interpretation of laboratory and field experiments aimed at determining the necessary protective threshold quantities of wood preservatives. It details the procedure followed in the soil-block tests at Bell Telephone Laboratories, Incorporated. Discussion of specific criticisms of the techniques involved and replies to these criticisms are included. The paper also presents for the first time a correlation of the results obtained from soil-block culture tests, outdoor exposure tests on stakes and on pole-diameter posts as well as pole line experience.

TABLE OF CONTENTS—Part II

Evaluation by Treated $\frac{3}{4}$ Inch Southern Pine Sapwood Stakes in Test Plots	427
Rating the Condition of the Stakes	427
Depreciation Curves for $\frac{3}{4}$ Inch Stakes	431
Estimating Threshold Retentions and Average Life	434
Evaluation by Treated Pole-Diameter Posts in Test Plots	443
Evaluation by Pole Test Lines and by Line Experience; Service Tests	449
Discussion	451
Density and Growth Rate	452
Size and Shape of the Test Blocks	452
Toluene as a Diluent for Creosote Treating Solutions	455
The Distribution of the Preservative in the Block	456
Heat Sterilization of the Treated Blocks	456
The Weathering of Creosote and Creosoted Wood	457
General Considerations; Creosote Fractions	458
Creosote Losses	461
Creosote Losses from Treated Blocks	469
Creosote Losses from Impregnated Filter Paper	473
An Interpretation of Creosote Losses	474
The Gross Characteristics of the Residual Creosotes in Soil-Block Tests of Weathered Blocks	479
The Evaluation of Greensalt	487
The Evaluation of Pentachlorophenol	488
Swedish Creosote Evaluation Tests	489



Shortening the Bioassay Test.....	490
Toughness or Impact Tests for Determining Preservative Effectiveness.....	491
Other Accelerated Bioassay Tests.....	493
Other Observations.....	494
Conclusions.....	497
Acknowledgments.....	498
Bibliography.....	499

EVALUATION BY TREATED $\frac{3}{4}$ -INCH SOUTHERN PINE SAPWOOD STAKES IN TEST PLOTS

Rating the Condition of the Stakes

One of the general and unavoidable difficulties in experiments involving exposure of small specimens in test plots is arriving at a measure of the inspector's judgment of the condition of the individual specimens at each inspection period. Bell Laboratories' investigators have used a system of numbers beginning with 10 as the highest, and running down in single steps to 0, to define the various gradations of destruction shown in the specimens as they pass from perfectly sound to the state of "failed" units. This system has been considered by some as slightly cumbersome; but it is a truly effective method of depreciation rating in a continuous series of inspections. In such a series any minor errors of judgment in one season can be corrected in the next. Slow depreciation can be recorded in the upper ratings until progressive destruction becomes clearly evident. Most observers of test plot experiments on treated wood specimens use a series of five numbers for five condition categories, about as follows:

- 10.0 Sound—no decay.
- 7.5 Surface soft—suspicious of decay.
- 5.0 Slight—positive decay.
- 2.5 Severe—deep decay.
- 0.0 Failed—almost complete loss of strength.

Some, like Rennerfelt⁸⁸ for example, use this system upside down, with 0 for no decay, and 10 for failure. The writer has proposed⁸⁵ the use of a new 5-number depreciation system, with the same definitions, based on the logarithms of the above 5 figures, and rounded off to 10, 9, 7, 4 and 0 respectively. This simplifies the Bell Telephone Laboratories' 11 division, 10-0, system while retaining the advantage of slow depreciation at first; and at the same time it avoids the sudden, and in the writer's opinion, unjustified drop from 10 to 7.5 in the arithmetic series for suspicious-of-decay specimens. In the following presentation and inter-

pretation of the behavior of some of the $\frac{3}{4}$ -inch stakes in the Gulfport test plot the recorded per cent condition of the stakes at any one inspection period has been translated into terms of the proposed 5-number log base system.

The stakes were carefully sawed and planed units, $\frac{3}{4}$ -inch square in cross-section.^{12, 69} Before treatment the stakes were selected so that they would represent the normal distribution of density in the material available. Table IX shows the analyses of the four different creosotes used to treat the stakes in respective 4- and 8-pound groups. Both empty-cell and full-cell treatments were employed. The full-cell treatments were made with toluene as a diluent in order to provide more uniform and lower controlled retentions in the treated specimens. The empty cell specimens were sorted after treatment to retain the middle group of retentions, with a view to eliminating as far as practicable some of the factors in the empty-cell treatment variation. All of the stakes were treated between March 11 and March 26, 1941, and they were all placed in the plot in the approximate period from April 8 to April 22, 1941. The distribution of retentions in the 8-pound stakes set out in the Gulfport plot are shown in Table X, along with data on average retention, standard deviation, and coefficient of variability.

TABLE IX — ANALYSES, WATER-FREE BASIS, OF FOUR CREOSOTES
USED IN TREATING $\frac{3}{4}$ -INCH SOUTHERN PINE
SAPWOOD STAKES
1941 series; Gulfport test plot

Creosote BTL No.....	5283	5286B	5286A	5285A
Specific gravity, 38/15.5°C.....	1.055	1.053	1.068	1.111
Distillation, per cent, cumulative				
To 210°C.....	2.4	4.6	5.1	0.5
210-235.....	13.2	22.4	20.3	4.2
235-270.....	38.1	49.3	41.1	18.8
270-300.....	51.1	60.5	50.0	28.7
300-315.....	58.4	65.6	53.5	33.1
315-355.....	81.1	80.7	67.3	53.2
Residue.....	18.8	18.6	32.5	46.7
Total.....	99.9	99.3	99.8	99.9
Sulph. res., gm/100 ml.....	3.4	1.6	0.7	0.7
Tar acids, gm/100 ml.....	7.8	5.7	4.0	4.0

TABLE X — DISTRIBUTION OF RETENTIONS,* LB/CU FT AT
TREATMENT, OF FOUR CREOSOTES, 8 LB EMPTY-CELL
(EC) AND FULL-CELL (FC) GROUPS

¾-inch southern pine sapwood stakes; 1941 series; Gulfport test plot.

Creosote, BTL No.	5283		5286B		5286A		5285A	
lb/cu ft	EC	FC	EC	FC	EC	FC	EC	FC
6.4	—	—	2	—	2	—	2	—
6.5	—	—	—	—	—	2	—	—
6.6	—	—	4	—	6	2	2	—
6.7	—	—	—	—	—	2	—	—
6.8	—	—	—	2	—	—	4	—
6.9	—	—	—	—	—	2	—	—
7.0	4	2	2	—	—	—	2	—
7.1	—	1	—	—	—	2	—	2
7.2	6	3	8	—	4	4	—	2
7.3	—	2	—	2	—	2	—	6
7.4	6	4	4	—	2	4	6	2
7.5	—	—	—	4	—	2	—	2
7.6	4	2	4	—	8	6	2	3
7.7	—	2	—	2	—	—	—	3
7.8	—	4	2	4	4	—	4	4
7.9	—	6	—	6	—	—	—	7
8.0	2	3	2	—	4	—	6	2
8.1	—	1	—	4	—	—	—	2
8.2	—	2	—	4	—	2	—	1
8.3	2	2	4	2	—	2	—	5
8.4	—	—	—	2	—	—	—	2
8.5	4	—	—	—	4	4	6	—
8.6	—	1	—	—	—	—	—	1
8.7	—	—	—	2	4	2	—	—
8.8	—	1	2	—	—	—	—	—
8.9	2	4	—	2	2	2	2	6
9.1	2	—	—	2	—	—	2	—
9.3	—	—	4	2	—	—	2	—
9.4	—	1	—	—	—	—	—	—
9.5	8	—	—	—	—	—	—	—
9.6	—	1	—	—	—	—	—	—
9.7	—	1	2	—	2	—	—	—
9.8	—	1	—	—	—	—	—	—
12.5	—	—	2	—	—	—	—	—

n

40	44	42	40	42	40	40	40	50
----	----	----	----	----	----	----	----	----

Average Retention lb/cu ft

8.12	8.02	7.97	8.07	7.76	7.59	7.79	7.91
------	------	------	------	------	------	------	------

Standard Deviation

0.913	0.723	1.288	0.514	0.839	0.694	0.758	0.523
-------	-------	-------	-------	-------	-------	-------	-------

Coefficient of Variation

11.24	9.01	16.16	6.37	10.81	9.14	9.73	6.61
-------	------	-------	------	-------	------	------	------

* All retentions were calculated from weights before and after treatment. The full cell (FC) treatments were made with toluene-creosote solutions.

TABLE XI — INSPECTION RATINGS

$\frac{3}{4}$ -inch southern pine sapwood stakes in test 6 and 7 years; 1941 series; Gulfport test plot.

Creosote BTL No.	Treat- ment	n	Average retention at treatment lb/cu ft	Test period years	Number and per cent of specimens rated					Average per cent condition
					10	9	7	4	0	
5283	EC	40	8.12	6	29 72.5	8 20.0	1 2.5	— —	2 5.0	92.2
				7	17 42.5	17 42.5	2 5.0	1 2.5	3 7.5	85.6
	FC	44	8.02	6	16 36.4	22 50.0	1 2.3	3 6.8	2 4.6	85.7
				7	2 4.6	19 43.2	14 31.8	3 6.8	6 13.6	68.4
5286B	EC	42	7.97	6	25 59.5	10 23.8	4 9.5	1 2.4	2 4.8	88.6
				7	9 21.4	14 33.3	10 23.8	4 9.5	5 11.9	72.0
	FC	40	8.07	6	10 25.0	27 67.5	1 2.5	1 2.5	1 2.5	88.5
				7	1 2.5	19 47.5	18 45.0	— —	2 5.0	76.8
5286A	EC	42	7.76	6	25 59.5	15 35.7	1 2.4	1 2.4	— —	96.5
				7	8 19.0	32 76.2	— —	— —	2 4.8	87.6
	FC	40	7.59	6	10 25.0	27 67.5	1 2.5	1 2.5	1 2.5	89.0
				7	8 20.0	20 50.0	8 20.0	1 2.5	3 7.5	80.0
5285A	EC	40	7.79	6	36 90.0	4 10.0	— —	— —	— —	99.0
				7	29 72.5	9 22.5	— —	2 5.0	— —	95.0
5285A	FC	50	7.91	6	35 70.0	12 24.0	— —	2 4.0	1 2.0	93.2
				7	17 34.0	23 46.0	4 8.0	2 4.0	4 8.0	82.6

Depreciation Curves for $\frac{3}{4}$ Inch Stakes

Under the conditions at Gulfport the depreciation curves for the creosotes — particularly the low residue oils — show an increased downward pitch at the 6th to 7th year of exposure. The relative proportion of the stakes rated respectively at 10, 9, 7, 4 and 0 at the 6- and 7-year inspections, are shown by number and per cent in Table XI. The change within the one year interval is particularly striking in the 10 and 9 columns. The rating and the distribution of retentions of creosote 5283, empty-cell treatment, at 6 and 7 years, respectively, are shown in Figs.

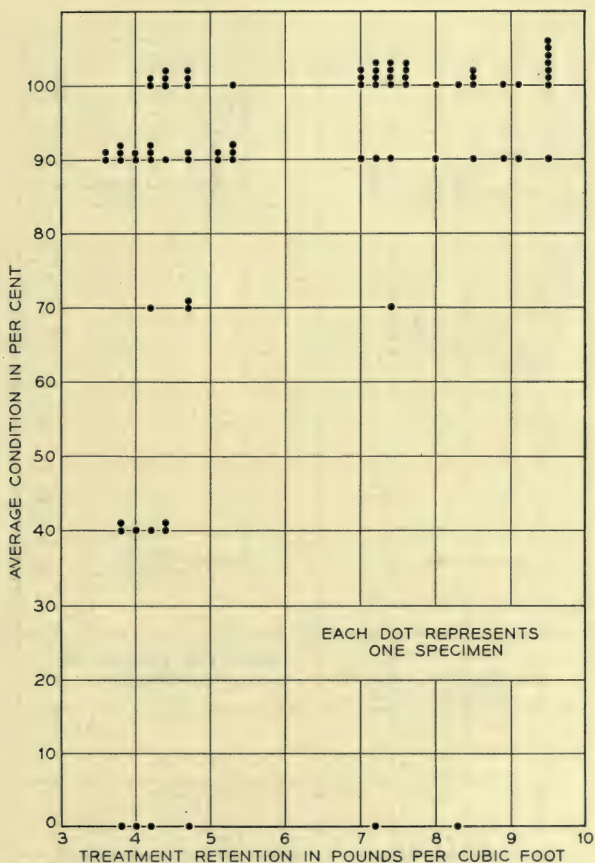


Fig. 15—Distribution of ratings in relation to retention by weight at treatment; creosote No. 5283, $\frac{3}{4}$ -inch southern pine sapwood stakes; 1941 series, empty-cell treatment; six years exposure; Gulfport test plot. See text and companion Figs. 16, 17 and 18.

15 and 16; and similar data for the full cell toluene dilute treatments are shown in Figs. 17 and 18, respectively. In spite of the attempted care in selection of the specimens and in treatment of the empty cell 8-pound group, the lot is heavily weighted (see Table X) toward the high retention end, e.g., by the 12 stakes in the 8.9, 9.1 and 9.5 groups. The presence of these relatively heavily treated stakes may explain in part the position of the depreciation curves for the 8-pound treatments shown in Fig. 19. On the other hand the empty cell 8-pound (nominal) stakes treated with

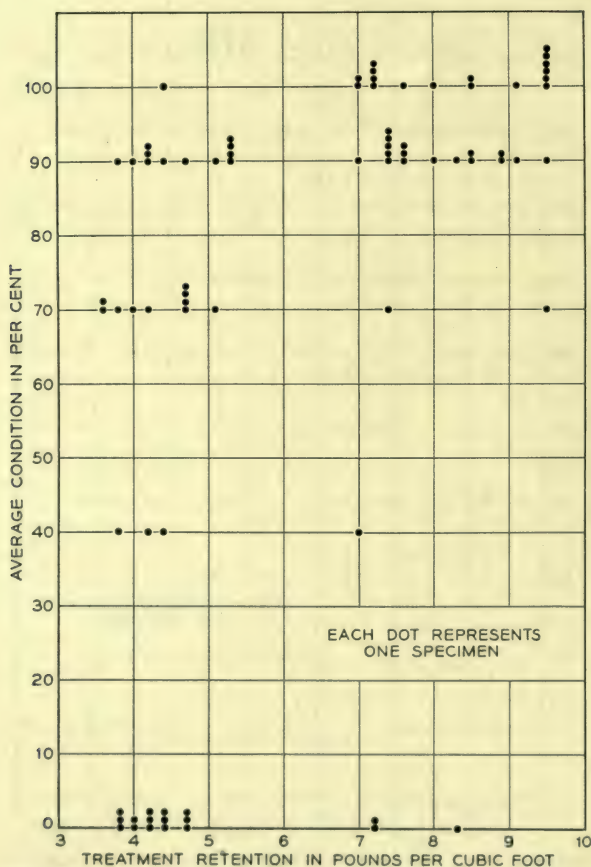


Fig. 16—Distribution of ratings in relation to retention by weight at treatment; creosote No. 5283, empty-cell treatment, seven years exposure.

creosote 5286B at 9.3, 9.7 and 12.5 pounds (see Table X) apparently have not operated to increase the average life of the group treated with this oil as much as appears to be the case in the group treated with creosote 5283. The difference in behavior at the 6-7 year interval of the stakes that were treated by a full cell process with treating solutions made by dissolving the creosote in toluene is even more marked than it is in the empty cell groups.

The average per cent condition of the stakes over the 9-year test period up to 1950 is shown in Table XII. Data are included on the number of stakes in each lot, and on the average treatment retention in

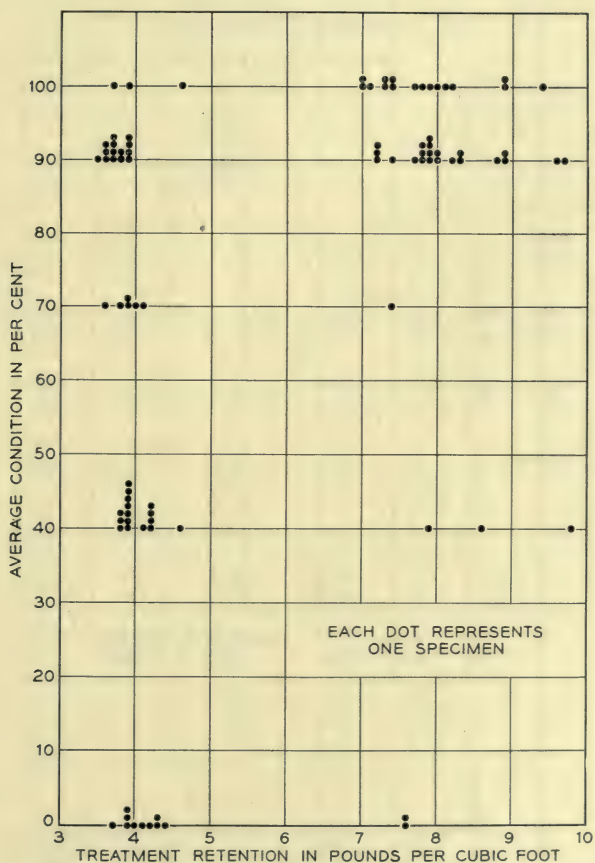


Fig. 17—Distribution of ratings in relation to retention by weight at treatment; creosote No. 5283, full-cell treatment (toluene dilution), six years exposure.

pounds per cubic foot for the respective groups. Two groups of stakes treated with greensalt $K^{74, 79}$ at 0.57 and 1.17 pounds per cubic foot, respectively, are included in the table for comparison.

Depreciation curves for the 4- and 8-pound groups for the four creosotes are shown in Figs. 19, 20, 21 and 22. A depreciation curve for greensalt K specimens treated with 1.17 pound per cubic foot is included in Fig. 19.

Estimating Threshold Retentions and Average Life

In presenting the following discussion of a theoretical approach to the estimation of threshold retentions and average life no attempt has been

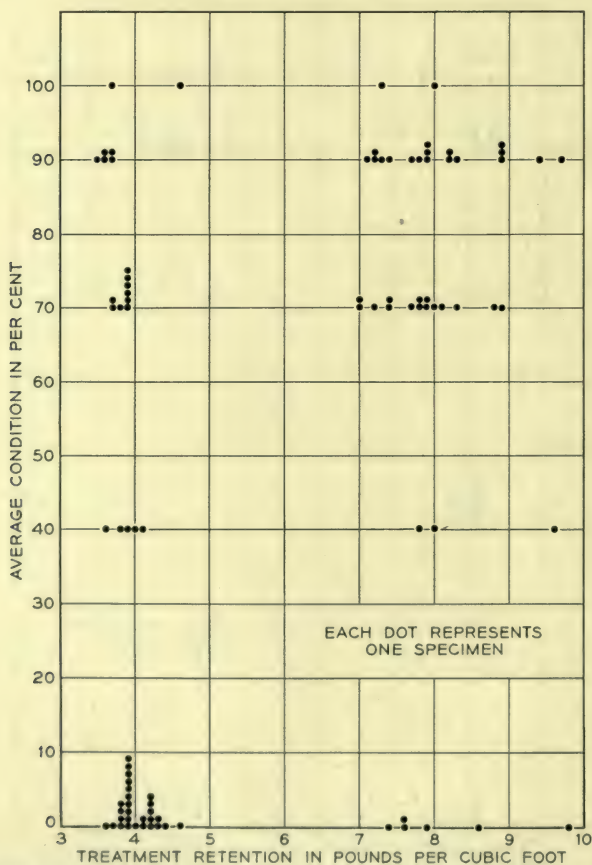


Fig. 18—Distribution of ratings in relation to retention by weight at treatment; creosote No. 5283, full-cell treatment (toluene dilution), seven years exposure.

made to separate the possible effects of the attack by different fungi or combinations of such fungi. The basic data are the figures reported by the inspectors of the stakes; and data on the actual organisms involved are very difficult — if not impossible — to obtain at the time of inspection. For the present purpose then, any differences in rate of decay by different organisms or in different parts of the test plot are all blanketed under the per cent condition averages.

TABLE XII — AVERAGE PER CENT CONDITION OF $\frac{3}{4}$ -INCH SOUTHERN PINE SAPWOOD STAKES TREATED WITH FOUR CREOSOTES, AND WITH GREENSALT K

9 years in test; 1951 series; Gulfport test plot.

Creosote BTL No.	Treatment	n	Average retention at treatment lb/cu ft	Years in test					
				1	3	6	7	8	9
				Average per cent condition					
5283	EC	40	4.38	100	97	74	50	32	13
	EC	40	8.12	100	99	92	86	74	55
	FC	49	3.93	99	86	54	30	16	8
	FC	44	8.02	99	96	86	68	54	31
5286B	EC	42	4.23	100	93	67	40	21	12
	EC	42	7.97	100	100	89	72	54	40
	FC	40	4.03	100	86	56	29	10	1
	FC	40	8.07	100	97	89	77	47	32
5286A	EC	42	3.93	100	96	76	57	42	28
	EC	42	7.76	100	100	97	88	82	72
	FC	40	4.08	100	93	84	64	44	16
	FC	40	7.59	100	96	89	80	72	53
5285A	EC	40	4.15	100	99	91	75	62	53
	EC	40	7.79	100	100	99	95	92	90
	FC	51	3.92	100	94	78	61	46	24
	FC	50	7.91	100	97	93	83	76	62
Greensalt K									
	FC	99	0.57*	100	98	84	80	72	67
	FC	100	1.17	100	99	96	88	86	79

* lb/cu ft of dry salt.

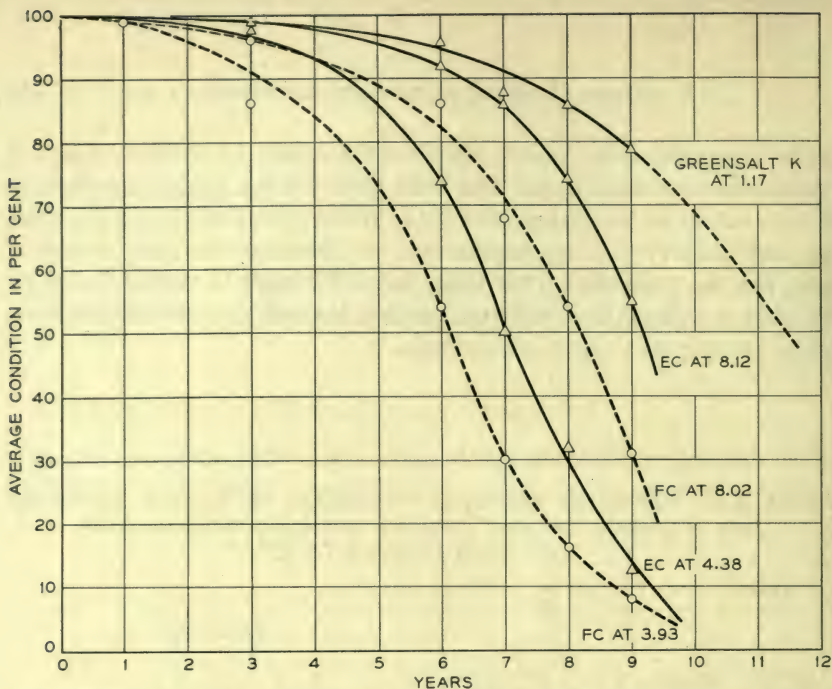


Fig. 19—Depreciation curves for $\frac{3}{4}$ -inch southern pine sapwood stakes treated with creosote, BTL No. 5283, empty-cell and full-cell (toluene dilution) processes, and with greensalt K; Gulfport test plot. See text, Tables XII-XIII, and companion Figs. 20, 21 and 22.

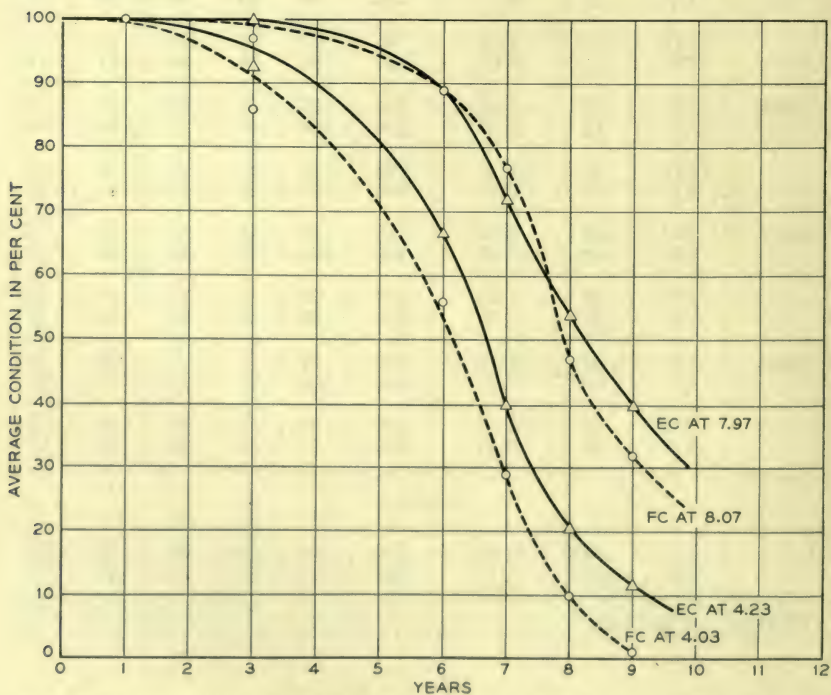


Fig. 20—Depreciation curves for $\frac{3}{4}$ -inch southern pine sapwood stakes treated with creosote, BTL No. 5286B.

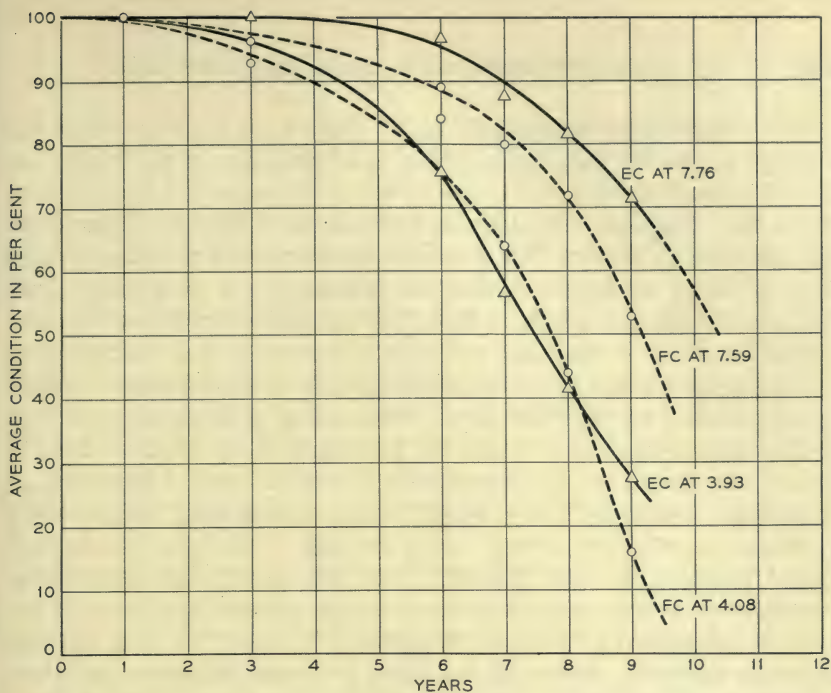


Fig. 21—Depreciation curves for $\frac{3}{4}$ -inch southern pine sapwood stakes treated with creosote, BTL 5286A.

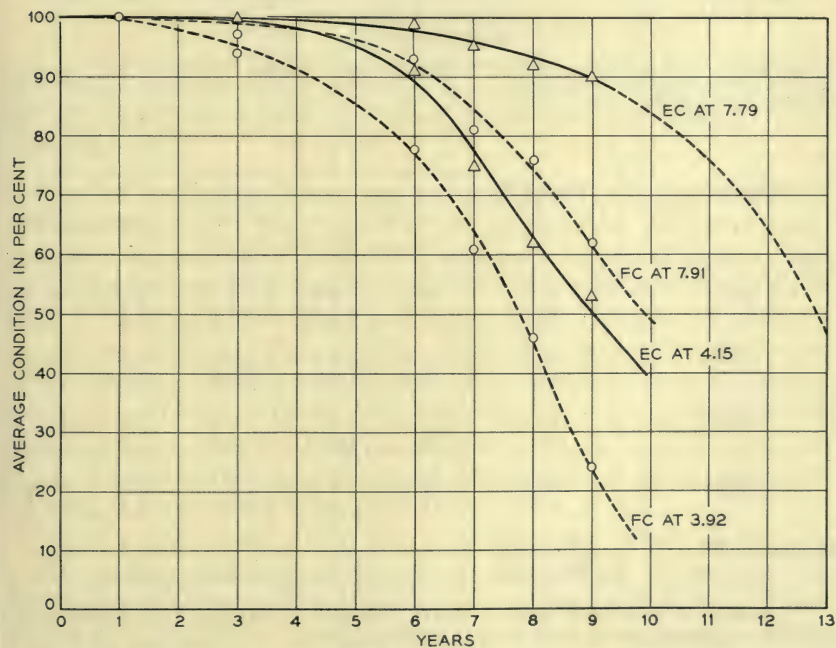


Fig. 22—Depreciation curves for $\frac{3}{4}$ -inch southern pine sapwood stakes treated with creosote, BTL-5285A.

The straight lines in Figs. 23, 24, 25 and 26 are drawn through the average points for per cent condition and retention of the respective 4- and 8-pound groups, and projected to intercept the 100 per cent condition line. The empty-cell data are represented by the solid line and the full-cell data by dashes. The threshold concentrations necessary to prevent all decay are estimated from the intersection of the gradient lines and the 100 per cent condition line. This method assumes that the relation of condition to treatment retention is linear at or near the threshold, providing the average condition points through which the lines are drawn are established by the logarithmic based rating system described. The method also resembles in a way the procedure of the Madison investigators who have used the intersection of straight lines drawn through operational losses and decay weight losses in estimating the threshold retention in creosoted blocks.³⁹ As in the case of the latter the method would probably be more precise if one had more points for average condition at average retention nearer the thresholds. At any event the system represented by Figs. 23-26 seems to be about the only one that indicates probable thresholds for these particular creosotes and these particular sets of data.

TABLE XIII — ESTIMATED THRESHOLD RETENTION AND AVERAGE LIFE

$\frac{3}{4}$ -inch southern pine sapwood stakes; (see Tables X-XII); 1941 series; Gulfport test plot.

Creosote No.	Treatment	Years in test			Average life-years*	
		3	6	7	"4 lb"	"8 lb"
		Estimated thresholds lb/cu ft*				
5283	EC	9.7	9.7	9.6	7.0	9.2
	FC	9.6	9.8	11.4	6.1	8.2
5286B	EC	8.0	9.8	11.3	6.7	8.3
	FC	9.2	9.4	10.0	6.2	7.9
5286A	EC	7.7	8.3	9.3	7.5	10.4
	FC	12.2	Indet.	11.9	7.8	9.2
5285A	EC	7.8	8.3	8.7	9.0	12.8
	FC	11.8	9.7	11.1	7.8	9.9
Greensalt K	FC (0.57)	—			10.5	
	FC (1.17)	—	1.4	2.1		11.4

* See text for additional data on method used in estimating the threshold retentions at treatment and the average life figures.

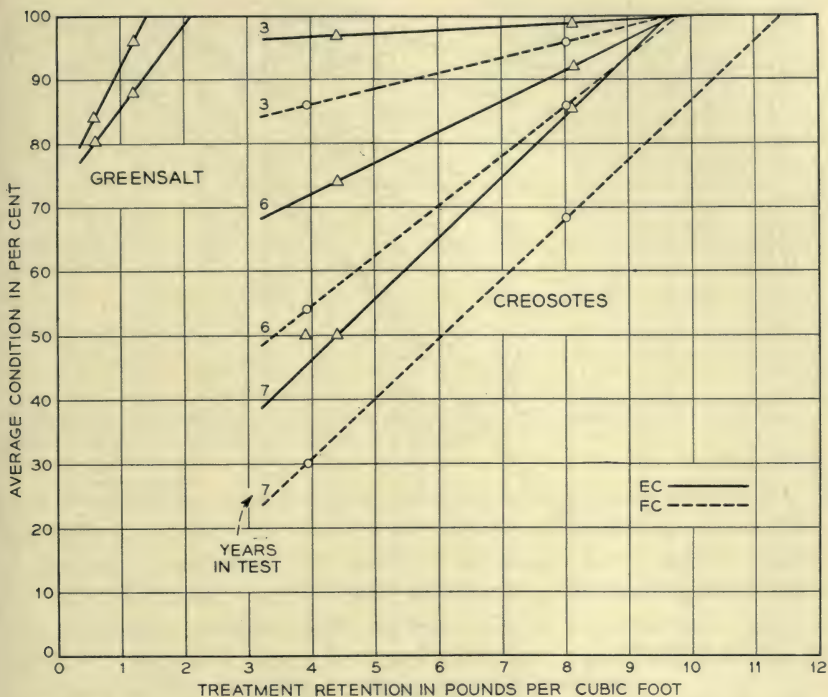


Fig. 23—Theoretical lines for estimating threshold retention for creosote No. 5283, in $\frac{3}{4}$ -inch southern pine sapwood stakes, empty-cell and full-cell (toluene dilution) treatments, and for greensalt K; Gulfport test plot. See text, Tables XI and XIII, and companion Figs. 24, 25 and 26.

The tendency for the 7-year lines to fall off to the right shows the effect of increasing decay in the 8-pound group. In the higher residue creosotes the earlier decay of the toluene dilute 8-pound treated specimens — that is, specimens that were treated below the threshold retention — tends to pull the lines so far down as to spoil their usefulness as tools for estimating thresholds. Obviously the slopes of the lines will be influenced by the depreciation rating of the 4-pound as well as the 8-pound groups. Furthermore it would appear that the utility of the specimens treated with toluene-creosote solutions for estimating thresholds does not extend much beyond the 6th year of exposure under conditions such as those prevailing at the Gulfport plot.

The estimated thresholds at the 3-, 6- and 7-year inspection periods, and the estimated average life values for the different groups are summarized in Table XIII. The average life is estimated from the intersection of the depreciation curves and the 50 per cent condition lines. There

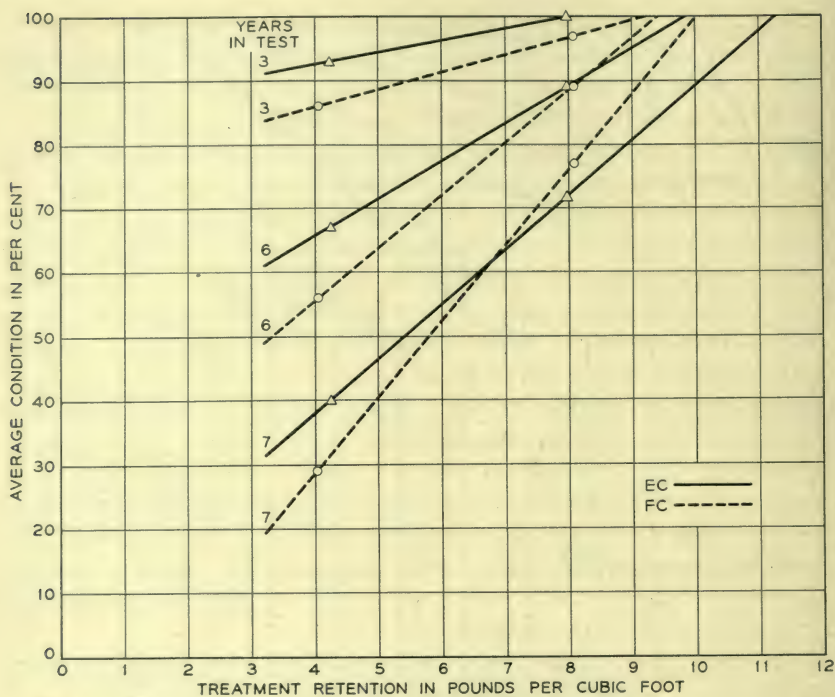


Fig. 24—Theoretical lines for estimating threshold retention for creosote No. 5286B; $\frac{3}{4}$ -inch southern pine sapwood stakes.

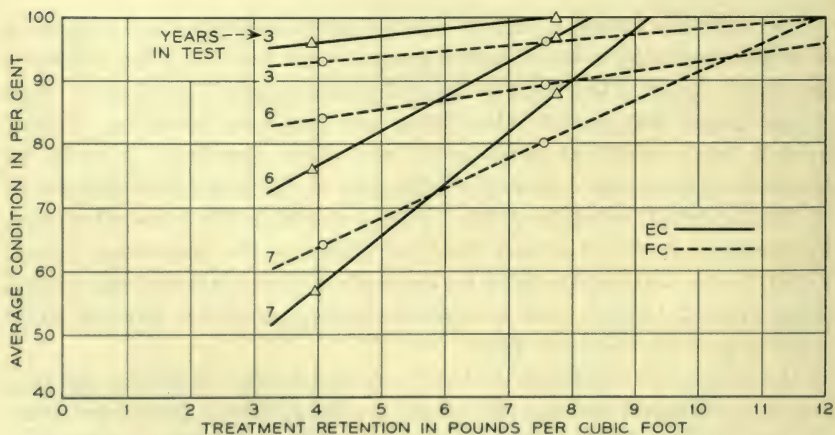


Fig. 25—Theoretical lines for estimating threshold retention for creosote No. 5286A; $\frac{3}{4}$ -inch southern pine sapwood stakes.

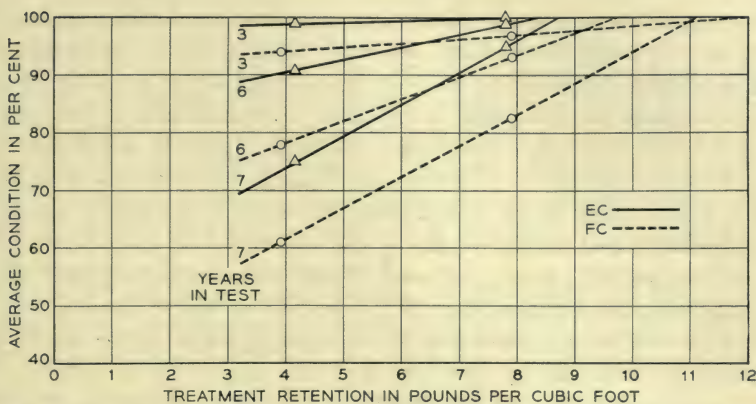


Fig. 26—Theoretical lines for estimating threshold retention for creosote No. 5285A; $\frac{3}{4}$ -inch southern pine sapwood stakes.

will inevitably be some difference of opinion as to which level to use. In the case of $\frac{3}{4}$ -inch treated stakes it is quite evident that the preservative is no longer functioning effectively if the stakes have reached a decay rating of 7 or less. In such small specimens it is questionable whether any purpose is served by allowing them to stay in the ground under the conditions at the Gulfport test plot until they practically fall over by being completely destroyed at the ground line.

Anyone who has worked with small test plot specimens will appreciate the many difficulties in the way of establishing standard procedure for determining the "failed" point or the end point of specimen life. On somewhat larger stakes Rennerfelt⁸⁸ has used a strength testing apparatus. To test the fitness of small poles in line some Associated Operating Companies have used a spring scale dynamometer which is slipped onto the base of a pike pole. In actual utility plant experience it is obvious that it is impossible to wait for the complete decay of the wood unit. Elaborate tables have been worked out as guides for pole line inspectors to let them know how far decay can go under given load conditions before a pole has to be removed from line. Generally speaking, such removal must occur at some period well in advance of the time that the specimen would have rotted clear through at the ground line. In the writer's opinion it might be preferable to estimate the average life for $\frac{3}{4}$ -inch stakes from the point where the depreciation curve passes downward through the 60 per cent condition line, leaving *all* the units in any given series in test until that time, except of course the stakes that may have actually rotted off earlier.

Study of Table XIII indicates clearly that both threshold and average

life estimates for any given set of small specimens will vary, depending *on the time at which the estimates are made*. Taking the 6-year inspection data as perhaps representing the best figures from which to make threshold and average life estimates of this kind, it appears that in the case of all of the four creosotes something more than 8 pounds of creosote per cubic foot was necessary to protect the $\frac{3}{4}$ -inch stake specimens against decay. There seems to be no material difference in the performance of creosotes 5283 and 5286B as far as the estimated thresholds are concerned. For both the empty-cell and full-cell treatments it appears to be somewhere in the neighborhood of 9.5 pounds or above. In the case of the higher residue oils 5286A and 5285A there appears to be a significant difference between ratings obtained from the empty-cell specimens and from the toluene dilute full-cell specimens. Estimates of average life from the 8-pound empty cell specimens appear to be significantly higher than estimates from the 8-pound full cell toluene-creosote specimens, except in the case of oil No. 5286B (Table XIII). The estimated thresholds for the full-cell toluene-creosote specimens lie within the same general magnitude as the retention at treatment thresholds found in the soil-block tests. (Cf. Tables XIII and XXXV).

How far one is justified in comparing straight 8-pound empty-cell treatments and 8-pound full-cell toluene-creosote treatments in $\frac{3}{4}$ -inch stakes is still not clear. It certainly cannot be done without taking into account the much greater variability in the retentions in empty cell stakes and the different but difficult to describe variations in distribution of the creosote from outside to inside, particularly in the case of empty cell treatments with high residue oil. Unless the empty-cell stakes are carefully selected within limited variations from the average retention, it has been found that the empty-cell coefficient of variability for retention in an 8-pound treatment, for example, may run as high as 35-40 per cent, against a coefficient of 8-10 per cent only for companion full-cell treatments with toluene-creosote solutions. The comparisons among the latter treatments appear to be more rational and fairer; and they may give a truer picture of effective threshold requirements.

The analysis of the stake test data here presented is intended merely to illustrate one set of procedures that may be used in the interpretation of small stake tests. The four sets of data are part of a much larger group of results that are being worked up for publication. Among the latter there are numerous lots indicating that truly protective thresholds of creosote for $\frac{3}{4}$ -inch stakes lie somewhere between 10 and 12 pounds per cubic foot, which is not far out of line with the estimates given above. Final analysis and publication will either confirm or modify such estimates.

EVALUATION BY TREATED POLE-DIAMETER POSTS IN TEST PLOTS

Reference was made in the introduction to the fact that Bell Telephone Laboratories' experience over the last quarter of a century in the evaluation of preservatives by the use of pole-diameter posts in the Gulfport test plot was reviewed by Lumsden⁷⁶ in April, 1952, before the American Wood-Preservers' Association. A supplementary analysis and interpretation of some of the same evidence is attempted in the following paragraphs.

Data for 185 of the posts, the time of placement and the number of years in test, and the general condition of the posts as of the 1950 inspection, are shown in Table XIV. The data for the individual posts of these

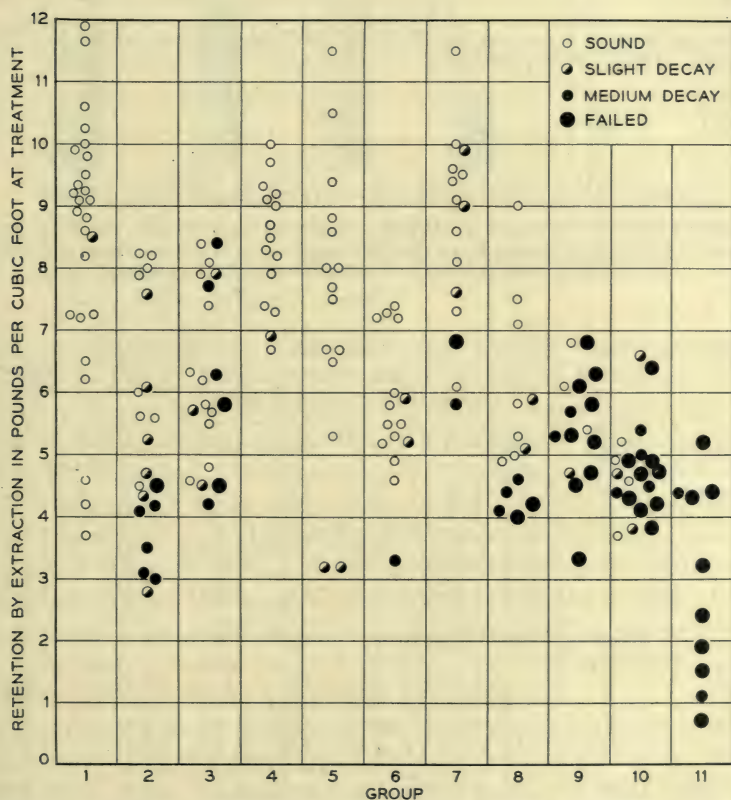


Fig. 27—The relation of creosote retention in lb/cu ft at treatment, by toluene extraction, and the rated condition of pole-diameter southern pine test posts, by post groups; 1950 inspection; Gulfport test plot. See text and Table XIV.

TABLE XIV — CREOSOTED SOUTHERN PINE TEST POSTS
Groups, years in test, and condition at 1950 inspection; Gulfport test plot.

Group*	Year placed	Years in test	n	Sound	Decaying		Failed
					Slight	Medium	
1	1931	19	26	25	1	—	—
2	1931	19	20	8	6	5	1
3	1931	19	20	11	3	4	2
4	1936	14	15	14	1	—	—
5	1935	15	15	13	2	—	—
6	1935	15	15	12	2	1	—
7	1935	15	15	10	3	1	1
8	1935	15	14	7	2	3	2
9	1935	15	15	3	1	2	9
10	1936	14	20	4	3	4	9
11	1935	15	10	—	—	2	8
Totals.....			185	107	24	22	32

* For analyses of creosotes, see Table II in Lumsden's report⁷⁶, Am. Wood Preservers' Assoc., Proc., 1952.

TABLE XV — CREOSOTED SOUTHERN PINE TEST POSTS

Relation between average treatment retentions by zones, by toluene extraction, and condition of the posts at the 1950 inspection, Gulfport test plot.

Condition	Group	n	Average retention at treatment lb/cu ft					
			Zones, inches				Remainder of treated sapwood	Whole cross section
			Outer ¼	Next ¼	Next ½	Next 1		
Sound	5-10	64	13.1	9.9	8.5	6.4	5.2	7.4
Decaying	5	2	9.7	7.3	4.9	—	3.2	4.2
	6	3	8.8	5.5	4.8	—	4.2	4.8
	7	4	13.1	11.6	11.8	11.1	6.5	8.1
	8	5	10.9	8.7	7.4	5.0	2.9	5.8
	9	3	12.7	9.2	7.0	2.9	2.9	5.2
	10	7	10.1	5.2	3.9	3.6	3.5	4.3
	Total	24	10.9	7.7	6.5	4.3	3.9	5.4
Failed	7	1	13.7	9.2	11.7	—	4.7	6.7
	8	2	9.2	7.4	6.8	—	3.6	5.1
	9	9	13.3	10.0	6.2	3.7	3.6	5.6
	10	9	9.2	4.9	4.0	3.2	3.3	4.2
	Total	21	11.1	7.4	5.6	3.5	3.5	5.0
	Overall	109	12.3	8.9	7.5	5.5	4.6	6.5

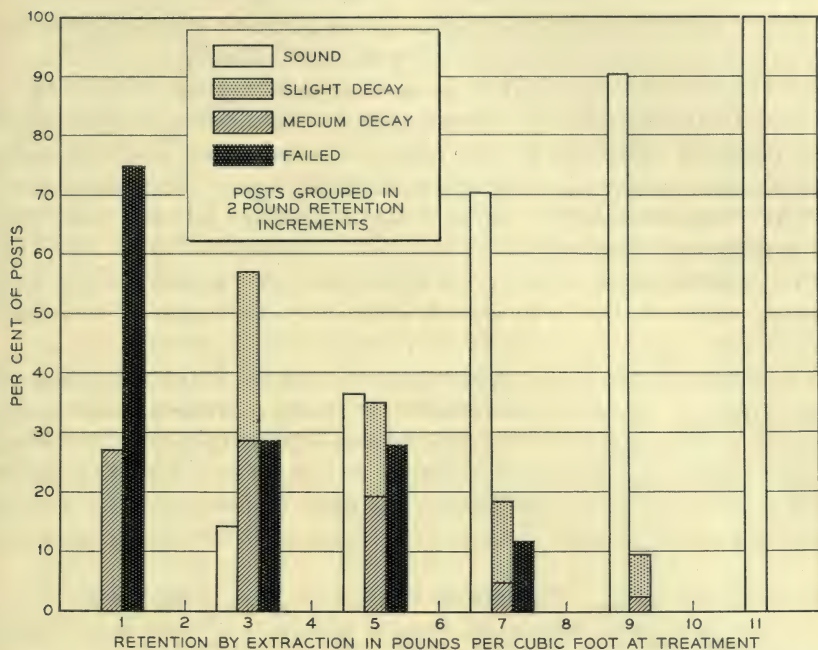


Fig. 28—The relation of creosote retention in lb/cu ft at treatment by toluene extraction, and the rated condition of pole-diameter test posts, by retention groups; 1950 inspection; Gulfport test plot. See text and Table XVI.

eleven groups are shown graphically in Fig. 27. The relation of retention by extraction at the time of treatment to the sound, decaying and failed specimens is clearly evident. All of the failures and the very great majority of the remaining decaying poles are below the 7-pound retention line. Only two cases of medium decay occur above the 7.5 pound line. It should be borne in mind particularly at this point that these retention levels represent the retentions found by extraction in the whole cross sections of the posts *as soon as practicable after* the posts were treated. The significance of these over-all retentions and of the calculated retentions in the outer one-inch layer of the posts will be discussed later. Since the data were calculated in terms of oven-dry weight and volume of the extracted wood the values may be a little high.

For Groups 5 to 10 inclusive data are available on the retention in zones, that is, in the outer quarter inch, the next quarter inch, the next half inch, the next one inch, and the remainder of the treated sapwood. These data were obtained by appropriate cutting of the samples into zones and pooling for extraction the parts that came from the same zones,

either in sectors cut from discs or in boring samples. A summary of the distribution of the retentions will be found in Table XV.

If the overall retention data for post groups 1-11 are distributed in 2-pound retention lots the evidence falls into the categories represented by Table XVI and Fig. 28. The inference is clear. One would expect a satisfactory service life if the treatment retention by extraction, based on the whole cross section, was at or above 8 pounds per cubic foot. This is perhaps over-simplification. The variations shown in Tables XIV and XV in the behavior of the posts in Groups 7 and 9, and possibly to a certain extent in Group 8, disturb what looks like reasonably straight reasoning.

In order to extend the reasoning and provide another method of interpretation, let it be assumed that the posts average 8 inches in diameter at the ground line, which is close to their actual size. Calculation of the pounds per cubic foot in the outer inch of such average posts is shown in Table XVII. The influence of posts in groups 7 and 9 on the averages is still evident. If there are differences in the efficiency of the

TABLE XVI — CREOSOTED SOUTHERN PINE TEST POSTS

Condition at 1950 inspection by retention groups; 14-19 years in test; Gulf-port test plot.

Retention* lb/cu ft at treatment	n	Condition of ground section, number and per cent			
		Sound	Decaying		Failed
			Slight	Medium	
1. 10.0-11.9	10 100.0	10 100.0	0 —	0 —	0 —
2. 8.0- 9.9	42 100.0	38 90.4	3 7.2	1 2.4	0 —
3. 6.0- 7.9	44 100.0	31 70.4	6 13.6	2 4.6	5 11.4
4. 4.0- 5.9	71 100.0	26 36.6	11 15.5	14 19.7	20 28.2
5. 2.0- 3.9	14 100.0	2 14.3	4 28.5	4 28.6	4 28.6
6. 0.0- 1.9	4 100.0	0 —	0 —	1 25.0	3 75.0
Totals	185 100.0	107 57.8	24 13.0	22 11.9	32 17.3

* The distribution into retention lots was made on the basis of both weight and extraction data, depending on information applicable to the eleven groups in Table XIV.

creosotes in these groups, the differences are not at present real and tangible because they are masked by other factors, among which distribution of the preservative, and retention and penetration variables seem most important. The over-all indications are that in general one should insist on something more than 8, and probably more than 9, pounds of creosote per cubic foot, at the time of treatment, *in the outer inch* of the ground section of a southern pine pole. This, in simple terms, is in line with the conclusions reached about threshold retention requirements from the laboratory tests of creosoted $\frac{3}{4}$ -inch cubes and from test plot results on $\frac{3}{4}$ -inch stakes.

It is much harder to rate treated test posts in terms of per cent condition than it is to rate small specimens. Fig. 28 must be considered therefore as a generalization from the plot inspection data. It must be interpreted with the help of Fig. 27 and the retention by zones data in Tables XV and XVII. The reader who is at all familiar with pole line service records will recognize how much more detailed information there is for these test posts than there generally is for the ordinary pole line. Yet it is practically impossible to draw up precise, unequivocal statements about the results in certain of the post series. The conclusions must be broad.

The structure of the wood in a pole, the distribution of moisture content, the variation in density of the wood from the outside to the inside annual rings, and the distribution of creosote from the outside toward the inside all go to make up a resultant that is obviously complicated. Some of these factors have been reduced to a schematic figure (Fig. 29)

TABLE XVII — CREOSOTED SOUTHERN PINE TEST POSTS

Relation between average treatment retention in outer inch of sapwood and condition of posts at the 1950 inspection; Gulfport test plot.

Group	Average retention at treatment in outer 1 inch of sapwood					
	Sound		Decaying		Failed	
	n	lb/cu ft	n	lb/cu ft	n	lb/cu ft
5-10	64	10.14*	—	—	—	—
5		—	2	7.13	—	—
6		—	3	6.09	—	—
7		—	4	12.11	1	11.61
8		—	5	8.71	2	7.90
9		—	3	9.38	9	9.15
10		—	7	5.95	9	5.95
Totals	64	10.14	24	8.03	21	7.59

* All calculations are made on the basis of an 8-inch diameter at the point of sampling.

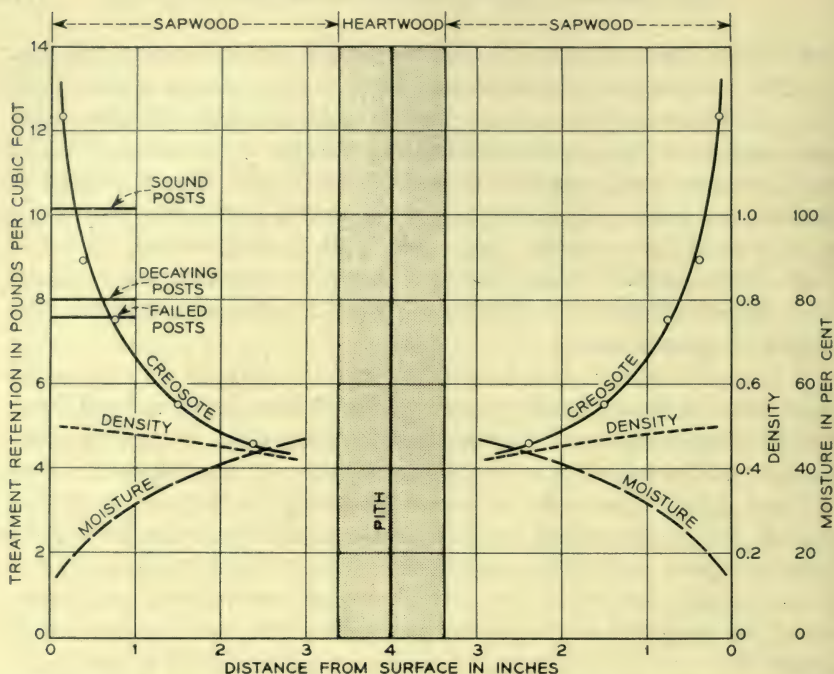


Fig. 29—Schematic diagram of a longitudinal section of a creosoted southern pine test post, showing average distribution of creosote in lb/cu ft; average density (oven-dry weight and volume as tested); average moisture content, oven-dry wood basis; and the retention levels in the outer inch of sound, decaying and failed specimens. See text.

representing a longitudinal section of an 8 inch diameter southern pine test post. The sound, decaying and failed retention levels calculated for the outer one inch (Table XVII) of such a diagrammatic post are also shown. When one considers all the variations in the wood itself and in the treatment, the site and environmental conditions where the pole is used, and the probability of the incidence of decay, the above general conclusions with respect to necessary retention seem reasonable and practicable. To refine and narrow the conclusions by repeating the post tests with the same creosotes — if one could get them — would certainly require the use of at least some posts with higher retentions and an obligatory assay of all of the treated specimens to be sure that the required retentions were actually in the wood in the right places. Apparently practical answers to questions about such required retentions can be answered very much more quickly by repeated series of soil-block tests in which many of the variables can be controlled.

EVALUATION BY POLE TEST LINES AND BY LINE EXPERIENCE; SERVICE TESTS

In 1932 a condensed report⁸⁴ of American Telephone and Telegraph Company experience with creosoted pole test lines, including brush, open-tank and pressure treatments, and covering northern (eastern) cedar, southern white cedar, chestnut and southern pine poles, ends with the following conclusions, among other generalizations:

"Because of the fact that many of the test specimens are still in service, conclusions can be reached only in the case of some of the less important problems whose solution has been sought. The possibility of extending the life of poles through preservative treatment is abundantly demonstrated, but the capabilities of the more effective processes of treatment (studies) can as yet only be estimated.

"... the indications are that in the cases not already affected, the beginning of decay attack will mainly be dependent upon changes in the quantity and the composition of the preservative retained in the individual poles."

This was only twenty years ago. Even in 1932 this service test report was more valuable as history than as a technical base for treatment specifications. The report reconfirmed the common knowledge that durable timbers like northern cedar and chestnut were better line units if properly butt-treated with creosote, and that adequate penetration and absorption of creosote were essential factors in the economy of pressure treatment of non-durable southern pine poles. However, by 1932

(a) The type of creosote used for treating the relatively large heart and small sapwood southern pine poles for the famous Washington-Norfolk and Montgomery-New Orleans lines was no longer available commercially;

(b) The type of virgin pine timber used was becoming scarcer and scarcer;

(c) The supply of commercial chestnut pole timber was just about completely exhausted;

(d) New and improved methods of butt treatment involving the use of machines for incising the ground section were being applied to northern cedar and to western cedar poles, and besides, the use of northern cedar was gradually shrinking to limited areas in telephone plant in the North-eastern and Lake States; and

(e) Because of vastly increased competition in the pole treating industry, and because the chestnut supply had failed, and because the excessive bleeding of creosote from the old style "12-pound" full-cell

southern pine poles — like those in the cited lines — made them unacceptable either as replacements for chestnut or for new construction in large sections of the northern and central telephone plant areas, and for basic reasons of economy, new specifications for creosoted southern pine poles were being issued. In preparing these specifications the Laboratories reduced the retention requirements to 8 pounds of creosote per cubic foot and provided for treatment by an empty cell process with a view to providing a clean pole, and set definite quality limits on penetration to insure an economic service life. These moves were made after careful experiment, and after analysis of 8-pound treatment results. The experience in the pole test lines had an indirect effect rather than a determining effect on the proposed new penetration requirements.

The Laboratories by 1932 had been operating the Gulfport test plot for 7 years and the Chester test plot for 3 years. Research and development programs on laboratory and test plot evaluation procedures^{123, 124, 129} were well under way. The aim of this broad program was to determine the necessary requirements for preservatives, for retention, for treatment and for penetration *before* the poles were placed in line; which is quite different from the philosophy of depending on service tests or records to reveal at some later date what was done wrong in the first place.

The reader should not get the idea that service tests can be dispensed with entirely. Material in service in the telephone plant is always under observation, casual or intensive, and the development of any obvious faults is corrected by the application of results of more and perhaps better laboratory experiments. In very many cases the faults are discovered in the laboratory or test plot before they are found in the field. However, one can never brush aside the insistent and determining effect of long and satisfactory field experience with treated wood.

For example, when the behavior of creosoted poles in line is good the service tests take on what seems to be the outstanding characteristic of their present function, namely that of a comforting confirmation of previous conclusions. The results of some of the Laboratories' analyses of the relation between penetration and decay in creosoted poles were published in July, 1936, in this Journal²⁹ and again in 1939.³ These results confirmed the previous actions involved in specifying an empty cell treatment and a retention of 8 pounds of creosote per cubic foot. The poles were clean, they were being accepted generally for line construction, and the incidence of infection and probable early failure seemed to be about in line with anticipation.

Emphasis in the last two papers cited was on penetration, which is easily determined, relatively speaking, and not on retention. The relation

of low retention to decay showed up in the Gulfport plot,⁷⁶ and it was explained by Laboratories' extraction and analysis of the creosote in the decaying or failed test stakes and test posts. One may say that such results confirmed suspicions; and they did, because creosoted poles in line — which were "related" to the decaying posts at Gulfport — were found on inspection to be behaving badly. It is believed that corrective measures in the way of supplemental ground line treatment were taken in time to give the poorly treated poles a reasonably satisfactory life.

In another striking set of circumstances, however, a number of cases of unprecedented premature failure of pine poles in line treated with a mixture of creosote and copper naphthenate petroleum revealed that the preservative solution had gone out of control and that in consequence the poles had not received the specified protective amounts of copper. Bell Telephone Laboratories' analyses of parts of the decaying poles showed that the decaying areas contained less than, and the sound areas more than, the Madison soil-block threshold of approximately 0.08 lb/cu ft of copper as metal.

These poles were immediately traced back to the supplier by their brand label. Present day pole treatment specifications in general require branding of each pole unit with symbols or code letters for the supplier's plant, the species of timber, the year of treatment, and the class and length of the pole. Each such pole becomes automatically a unit in a sort of universal service test in Bell System pole lines.

Service tests and service records are a significant part of the over-all process of evaluating wood preservatives, but insistence on service records as the most important criterion simply perpetuates the reputation of established preservatives and forever blocks or seriously impedes the development of new and promising materials. In the very nature of the case the results of service tests can be stated in the form of broad generalizations only. The truly technical approach must be made through better methods of measuring the effectiveness of preservative materials by accelerated field tests on a sufficiently large number of small stakes or by controlled laboratory experiments such as the soil-block tests.

DISCUSSION

There are a number of things in connection with the soil-block test procedure, its interpretation and the correlation of its results with the results of other evaluation methods that require further discussion. For example, questions have been raised vigorously on such matters as:

- (a) The effect of variation in growth rate and density of the wood;

- (b) The size and shape of the test block;
- (c) The use of toluene or any other diluent with creosote to get low retention;
- (d) The distribution of the preservative in the block;
- (e) The practice of heat sterilization of the creosoted blocks;
- (f) The whole philosophy of the weathering procedure; and
- (g) The methods of control assay, such as weight, creosote extraction, analysis of extracted creosote, and lime fusion chloride determinations as applied to pentachlorophenol.

None of these questions can be simply brushed aside, in spite of the fact that some of the points raised have a little of the nature of quasi-technical road blocks that may temporarily slow the approach to *any* standard laboratory test for creosote. Experimental work now under way at Madison⁴¹ and our own laboratories at Murray Hill will help answer some of the questions. The following discussion will at least explain the nature of the problems involved.

Density and Growth Rate

Data on the relation of block density and absorption at treatment have already been presented (Tables I, IV, V and VI). It is not yet evident that either density or growth rate has any material effect on the treatment retention thresholds for creosote when the tests are run on outdoor (or equally depleted) weathered blocks, and when the steps in the gradient retentions are properly spaced. When some present experiments are finished the matter may be resolved more conclusively.

Size and Shape of the Test Blocks

Some critics of the soil-block test have objected rather strenuously to the $\frac{3}{4}$ -inch cube because the two transverse faces are so close to each other. The inference is that either (a) the preservative, and in this case creosote is usually meant, is lost too rapidly through the end grain of the wood or (b) that in the evaporation process, or in the course of the weathering procedures, the preservative may be concentrated on the transverse faces. The validity of these contentions is under investigation. However, the importance of block size and shape can be greatly exaggerated.

One need not assume, in order to plan for comparative tests of oil preservatives, that a block must or can be cut so that creosote will be lost from it at exactly the same rate or manner as creosote may be lost radially from the concentric annual rings of a post or pole. Block shape

has varied almost as much as the test procedure employed. In a laboratory working plan for a study of the "Efficiency of Various Wood Preservatives," with the subtitle "The Efficiency of Various Wood Preservatives in Resisting the Attack of Fungi in Pure Cultures," signed by C. J. Humphrey, dated July 10, 1913, and approved by Howard F. Weiss, at that time director of the young U. S. Forest Products Laboratory, Humphrey proposed the use of eastern hemlock heartwood, cut into pieces measuring $1\frac{1}{4} \times 1\frac{1}{4} \times 2$ inches, which were to be treated and quartered longitudinally "*to reduce the size sufficiently to allow their introduction into the (Erlenmeyer) flask.*" (Writer's italics.)

Untreated wood blocks were to be used as culture media, and thirty days after inoculation the test pieces were to be put into the flasks and shaken up with the inoculated blocks. The test fungus was to be *Lentinus lepideus*. The culture period was to be 9 months, with culture blocks and mycelium "kept in a condition moist and warm enough for active growth" throughout the test period. Coincidentally, Humphrey⁶¹ and his colleagues were developing the broad foundation for the Petri dish toxicity test. The proposed wood block tests never reached a really satisfactory experimental level for treated wood, although they were employed for comparative natural durability tests.⁵⁷

Breazzano^{20, 21} in the same year experimented with blocks of beech wood cut from treated ties and measuring about $9 \times 2 \times 1$ cm. As has been pointed out earlier in this paper he recommended blocks measuring $4 \times 2 \times 1$ cm as a tentative standard, and later, in 1922, blocks $4 \times 4 \times 2$ cm. were accepted as standard in Italy.²² In the case of the latter the larger faces were to be transverse faces, cut across the grain.

Howe⁵⁶ reports tests of small sticks (blocks) of southern pine, measuring $\frac{3}{8} \times \frac{3}{8} \times 6$ inches, that were treated with salt preservatives and later inserted in "8-inch sterilized test tubes containing about 10 cc of standard malt agar." He also used sets of four small sticks in Petri dishes, placing them on 10 cc of nutrient agar medium that covered the bottom of the dishes. The fungus used was called *Fomes annosus*. To supplement these tests he mixed ground-up treated wood in different concentrations with agar media, and tested his mixes against this same *Fomes annosus* and a number of other wood-destroying fungi.

Howe and Curtin³² were working on a broad plan aimed at correlating laboratory tests with test plot tests made, e.g., at Matawan, N. J., and with experience in line.

Snell¹⁰⁷ argued that one might obtain good growth by placing thin blocks of wood in tubes with agar, but that the growth of the test fungus on the agar and on the wood might be influenced by diffusion of the

preservative into the agar medium. He proposed the use of thin plaques of wood, measuring $\frac{1}{8} \times 3 \times 3$ inches. These could be soaked to any required concentration of preservative and then tested by placing them over wet filter paper in Petri dishes. The test specimen would be supported just above the wet filter paper on sterilized wood strips. Inoculation was to be by the simple process of placing a small square of agar plus the growing test fungus directly on the upper surface of the test "block." Similarity to natural conditions and the control of moisture content of the wood were considered to be decided advantages for the method.

Cislak (see discussion of Snell and Shipley¹⁰⁸) used similar plaques of wood measuring 4 x 4 inches and about $\frac{1}{28}$ -inch thick for experiments on evaporation and permanency of creosote.

Rhodes, Roche and Gillander⁹⁰ used blocks measuring $\frac{1}{2} \times \frac{1}{2} \times 3$ inches.

The European standard^{33, 45} block measures 5 x 2.5 x 1.5 cm, with the long axis in the direction of the grain.

Schulze, Theden and Starfinger, in addition to the standard block, used "half" blocks, i.e., blocks measuring 5 x 2.5 x 0.75 cm. All factors considered, they do not regard the thinner block as an advantage, and they have held to the standard size.^{54 (1)}

Lutz⁷⁷ in 1935 suggested the use of 2 x 2 x 5 cm. blocks, with the long sides dressed parallel to the fibers of the wood. He also used blocks measuring 1 x 1 x 5 cm.

Alliot² favored blocks measuring 5.0 x 1.0 x 0.5 cm. in the longitudinal, tangential and radial directions, respectively, for a French standard test.

In his recent tests^{51, 52} Harrow has used $1\frac{1}{2} \times 1\frac{5}{16} \times \frac{7}{8}$ inch blocks.

Sedziak¹⁰⁶ uses $\frac{3}{4}$ -inch cubes cut from $\frac{3}{4}$ -inch stakes *after treatment*.

The National Wood Manufacturers' Association¹⁰⁹ standard block size is 1.25 inches on the radial surface, 1.75 inches on the tangential surface and 0.25 inch thick, i.e., in the longitudinal direction of the grain.

Various size blocks have been used at Madison in soil-block tests of natural durability, but two sizes only have been employed commonly since 1944 in the above mentioned soil-block tests and agar-block tests. The soil-block is the $\frac{3}{4}$ -inch cube, generally drilled with a $\frac{1}{8}$ -inch hole in the center of a tangential face. The agar-block was cut with two broad transverse surfaces measuring $\frac{3}{4} \times 1\frac{1}{2}$ inches, and with a distance along the grain of only $\frac{3}{8}$ inch; and it was not drilled.⁴¹ The Madison agar-block is basically the same sort of a block as the one described in the previous paragraph, and it resembles the Breazzano blocks²¹ as far as maximum transverse surface exposure is concerned.

The list is incomplete, but it serves to illustrate one important point about laboratory evaluation tests, and that is the inevitable variation that creeps naturally into explorative research. In the writer's experience and opinion much of the variation in block size has been the result of a sort of forced adaptation, on the part of the investigator, to the size and shape of his laboratory glassware, coupled with certain practical problems of block procurement and manufacture. Obviously it is easy to use thin sticks in test tubes, thin sticks or plaques in Petri dishes, and relatively flat blocks in Kolle flasks. The $\frac{3}{4}$ -inch cubes, which in essence, as has been stated before, are simply sections of the $\frac{3}{4}$ -inch stake, handle easily in the soil-block cultures.

Criticism of the shape of the $\frac{3}{4}$ -inch cube did not become pointed until after the publication of the first papers^{95, 35} on the Laboratories' cooperative work with the Madison laboratory. It is argued — as mentioned before — that the evaporation of creosote from such blocks is unfairly rapid. However, the losses reported by Rhodes et al,⁸⁹ which will be discussed later, indicate that separation of the transverse faces will not prevent creosote evaporation. Assuming properly calculated gradient retentions, and the use of weathered blocks, it does not appear likely that the shape of the blocks — if kept constant within any given comparative test series — will affect the location of the treatment threshold retention.

Toluene as a Diluent for Creosote Treating Solutions

This subject is most controversial in this country. The use of acetone or chloroform has been widely accepted in Europe, but Schulze and Becker¹⁰⁴ warn that the use of any diluent may change the rate of evaporation of given creosote fractions, and may affect the rate of evaporation of whole creosote. However, there are points in the debate which can be stressed, namely:

(a) The treatment of test blocks to low and uniform retentions without a diluent is extremely difficult, if not practically impossible;

(b) The rate of loss of creosote by evaporation, and any change in the character of that loss that may result from toluene dilution might be evident in freshly treated blocks but not in weathered blocks; and

(c) The volatile fractions of undiluted creosote are lost fairly rapidly from small saplings¹²⁵ and from test blocks; and it is assumed that the use of toluene does not cause loss in the fraction above 355°C.

Discussion of Item (c) will be resumed in the section on weathering. In view of the perplexing character of this toluene dilution question steps are being taken to find out what happens. In the meantime the toluene diluent which has not been found to exert any measurable toxic effect in

the soil-block tests at Madison⁴¹ must be used to secure gradient retentions down to and below the threshold level.

The Distribution of the Preservative in the Block

The opinion has been expressed by a number of investigators^{14, 22, 23, 100} that it is difficult to secure uniform distribution of the preservative in the test block, and that evaporation would cause concentration of certain preservatives on the block surface, where the toxic material might influence the behavior of the blocks in the test cultures. If such a concentration occurs the result would be to fix the threshold at lower over-all treatment retention than it would be if the preservative were not concentrated, for example, at the transverse faces. One is likely to agree that uniform distribution of some creosotes might be difficult without the use of a diluent, particularly in the low retention groups. Rhodes et al⁸⁹ were apparently satisfied that they got a fairly good distribution with a 16.7-pound retention. Experience at Madison and at Murray Hill indicates that the blocks are saturated with the toluene-creosote solution or with the toluene-penta-petroleum solution; and this confirms the ideas of Schulze, Theden and Starfinger^{54 (1)} on the way oven-dried test blocks take up the treating solutions.

Data on the distribution of residual creosote in weathered blocks is presented in Table XXX, and discussed in the section on weathering. Preliminary assay of blocks treated with penta-petroleum, (a) just after treatment, (b) after weathering and (c) after testing show that the pentachlorophenol concentration is slightly *lower* at the transverse faces than it is in the middle section of the blocks. The same seems to hold true for the copper metal in blocks treated with copper naphthenate in toluene-petroleum.

It should be pointed out here that preservatives are rather fortunately distributed in treated wood in such a way that the outer fibers or annual rings normally contain much higher concentrations of the toxic material than is found in the inner fibers. (See Tables XV and XVII and Fig. 29).

Heat Sterilization of the Treated Blocks

In pure culture test experiments some form of sterilization must be used to avoid contamination by other organisms than the test fungi. Most frequently such sterilization, for the minimum necessary time, is accomplished by flashing the treated block through a hot flame or by steaming at 100°C and atmospheric pressure. Either of these procedures

might cause a measurable loss of volatile creosote fractions from freshly treated blocks; but such losses appear to be negligible in the case of weathered blocks. At Bell Telephone Laboratories control data on these possible losses are being determined by extraction of control blocks after the sterilization phase. Such blocks are run through all the steps in the bioassay procedure up to planting in the soil-cultures. It will be recognized by anyone familiar with pressure treating methods that the 100°C temperature, usually held for 15 minutes only, represents a much gentler set of conditions than the after-treatment steaming for several hours at 240–259°F which is permitted in many specifications for creosoted poles.

The Weathering of Creosote and Creosoted Wood

Creosote is a remarkably good wood preservative, and nothing in the following paragraphs is intended to detract from its reputation in that respect. Sometimes the creosote oozes or bleeds from the surfaces of treated units such as poles and crossarms, especially on hot, sunny days; and when such bleeding occurs the treated material is unsatisfactory for use in many parts of the telephone plant. In order to prevent bleeding difficulties and the consequent unhappy employee and public relations that result, the retention requirements have been held down to the commercial standard level of 8 pounds per cubic foot for southern pine poles, and the residue above 355°C limitations on the creosote have been kept in actual practice at 25 per cent or below except when war or post-war emergencies have interfered.

The creosotes vary in the proportion of readily volatile materials they contain, and these materials are lost from creosoted wood — largely by evaporation — under many different use and exposure conditions. The general facts about such losses have been reported over and over again since the turn of the century. The significance of such losses is still not broadly understood or appreciated. Their possible bearing on the weathering procedure to be used in the soil-block tests is of fundamental importance. Paraphrasing the quotation from Schmitz⁹⁹ cited in an earlier paragraph: It is really necessary to know how much creosote to inject into wood to allow for loss by volatility and to insure a residual of the preservative, remaining in sufficient amount, to protect the wood for an economical service period. Bell Laboratories has been deeply concerned with the question whether it is practicable under commercial conditions to specify enough retention at treatment to provide the necessary protective residual and at the same time require clean, satisfactory, treated poles and crossarms for Operating Company use.

General Considerations; Creosote Fractions

Before presenting definite evidence of creosote losses this seems as good a place as any to refer briefly to investigations that have been aimed at discovering what components, volatile or relatively stable, give creosotes their properties of toxicity and permanence. Three articles by Martin,⁸⁰ Rhodes⁹⁰ and Mayfield⁸¹ are the latest American papers covering the general subject, the two former dealing with the technology of hydrocarbons and creosotes and creosote production and the latter reviewing the results of tests of whole creosotes and creosote fractions.

Teesdale's short report¹¹³ in 1911 is one of the earlier records in this country of experiments aimed at determining the loss of creosote fractions from treated wood. He used a creosote with 49 per cent distilling below 250°C and a residue above 320°C of 28 per cent, which would have been about 20 per cent at 355°C. This oil was fractionated into 5 parts, I, to 205°C; II, 205–250°C; III, 250–295°C; IV, 295–320°C; and the residue above 320°C. These five fractions and a sample of creosote with a similar distillation range were used to treat air-seasoned pieces of *Pinus taeda*, mostly sapwood, cut 2 feet long from 5 to 6-inch diameter peeled posts. The retentions were about 18 pounds for the numbered fractions, 15 pounds for the residue, and 21 pounds for the whole creosote.

The treated pieces were open-piled in the laboratory for two months, with temperatures running from 60 to 80°F. At the end of that period the per cent losses of the five fractions and the creosote were respectively 34.7, 21.3, 15.9, 6.2, 4.0 and 5.4. The results were in line with expectations. The test period was short, and there were no outdoor weathering factors. Teesdale notes that the loss of the whole creosote was about the same as the losses in the two higher fractions, III and IV; and that a proportionately composited sample of the five fractions lost at the rate of fraction III, the total at the end of the two-month period being 15.8 per cent or about *three times* as great as for the whole creosote.

Loseby and Krogh⁷³ reported in 1944 on outdoor weathering tests of creosoted wood blocks; and they compared the weight losses in the blocks with evaporation losses from open Petri dishes. The creosote used was a relatively low residue oil produced at Pretoria. The residue above 355°C was 19.95 per cent, the specific gravity at 38/20°C was 1.088, and the amount distilling to 235°C was only 4 per cent. The test blocks were planed pieces of light weight *Pinus insignis* measuring 6 x 1½ x 1½ inches. The per cent moisture at the time of treatment was 10.3 per cent.

The creosote was fractionated into four parts; Fraction I, up to 270°C; Fraction II, 270–315°C; Fraction III, 315–355°C; and, Fraction IV, the

residue above 355°C. The blocks were very heavily treated by a hot and cold soaking process with the straight creosote and each of the four fractions to the following average retentions, respectively: 52.2, 45.6, 50.3, 50.2 and 23.7 pounds per cubic foot. These high retentions place the experiments out of line with most of the others cited in this paper; but the South African tests are unique in that they supply evidence on the rate of creosote losses from such high retentions.

The relative order of losses of the materials, with the one having the highest losses first, was the same for the blocks that were hung on wire outdoors and for the open Petri dish samples, namely; Fraction I, II, whole creosote, Fraction III, and the residue. The latter actually showed no significant loss in either block or dish tests, and in the blocks there was a slight increase, possibly referable to oxidation, of a maximum of 2.2 per cent at the end of the three-year period. This gain was gone by the end of the 5½-year test. Fraction III was lost more rapidly from the blocks than from the *Petri dishes*; the reverse was markedly evident for Fractions I and II; whereas the pattern for the whole creosote was very similar after the first year. The rounded figures for losses from blocks treated with the whole creosote, at the end of 1, 2, 3 and 5½ years outdoor exposure, were 36, 42, 44 and 47 per cent; and the losses from the open dishes for the same periods were 34, 40, 43 and 48 per cent respectively. One may conclude from these tests and conditions that a loss level of about 50 per cent would have been reached in about six years in blocks treated to a reported 50-pound per cubic foot retention with an undiluted creosote.

Most of the laboratory toxicity tests on fractions have been run by the agar or the agar-block method.^{54 (1), 104, 11, 101, 47} The results obtained by Schulze and Becker¹⁰⁴ are cited by Mayfield, who includes as his Fig. 1 a copy of the summary curves prepared by the Berlin investigators. The interested reader can profitably use the rather full excerpts of tabulated results of other investigators given in Mayfield's paper as an introduction to the difficulties of testing creosotes and of interpreting the results of such tests. Some of the main controversial points are brought out by Peters, Krieg and Pflug in 1937⁸³ who challenge the results of Petri dish agar toxicity tests with results of the German agar-block tests — one of the first such broad comparisons to be made. Broekhuizen²⁵ published his findings the same year in a comprehensive paper covering agar-block tests on creosotes and creosote relatives and creosote fractions. He discusses his results with different preservatives in relation to their toxic properties, their protective and preservative qualities, and the *permanence* of such qualities, and the bearing of these qualities on practical

wood preserving procedures. His discussion of the importance of "weathering" tests and of his own results with such tests make up one of the best, if not the best consideration of this important phase of evaluation tests that had appeared up to that time.

Numerous references to weathering techniques will be found in van Groenou, Rischen and van den Berge;¹¹⁸ and van Groenou's own paper¹¹⁷ in 1940 is an excellent short review of previous work, giving his views of the pros and cons of different procedures with emphasis on the essential nature of some test to determine what changes are likely to take place in a preservative as a result of leaching or evaporation.

The following examples will be interesting as illustrations of the different techniques that have been employed in testing the toxicity, or the potential preservative value of creosotes. F. H. Rhodes and Gardner⁹¹ in 1930 determined evaporation losses of creosote fractions from thin pads of dried ground Sitka spruce pulp. The whole creosote and the fractions were introduced in an ether solution. The impregnated pads were placed on top of agar cultures in Petri dishes. The test fungus was the usual *Fomes annosus*, more recently called simply Madison 517. They state that:

"It was found that under these conditions (the Petri dish covers loosely fitting — not water sealed) the more volatile preservatives vaporized from the test specimen, so that at the end of the month only a relatively small portion of the fungicide remained in the pulp."

They were using a domestic creosote with a specific gravity of 1.065 at 38/15.5°C and a residue above 355°C of 21 per cent. They tested fractions of the dead oil from which the tar acids and tar bases were removed, and fractions of the tar bases and the tar acids themselves. They determined percentage losses by evaporation for all lots by letting the treated pulp disks remain in covered Petri dishes for one month at 25°C. All of these reported losses for the dead oil occurred in fractions boiling below 316°C, for the tar acids in fractions with the same upper limit, and for the tar bases in the fractions boiling below 308°C. The losses varied inversely as the boiling range, the greater being in the low-boiling fractions, as was to be expected. The toxic limits for *Fomes annosus* that they determined showed a gradual increase as the boiling point increased, i.e., in agreement with other workers with agar tests they found the lower boiling fractions the most toxic. Their bibliography, with one or two exceptions, covers American articles only.

Rhodes and Erickson,⁹² continuing the same general technique, but substituting mechanical pine pulp for spruce pulp, showed that much

higher quantities of respective creosote fractions were required to kill *Fomes annosus* in the pulp cultures than in Petri dish agar cultures. They concluded from their experiments that "no one compound in coal-tar creosote is primarily responsible for its preservative power. . . . The fractions from water-gas tar oil are much less effective as preservatives than are those from coal-tar creosote oil. The chlorine derivatives of phenols and creosote and of naphthalene are more toxic to fungi than are the compounds from which they are obtained."

In 1933 Flerov and Popov⁴⁸ tested fractions of two creosotes by their soil-block method and compared the results with tests of the fractions emulsified in agar, following American Petri dish procedure in general. They found (English translation by Hildegard Kipp, Forest Products Laboratory):

"On wood the toxicity of the heavy fractions is considerably higher (compared to that of the lower fractions) and the most toxic fractions are those from 315 to 375°C. . . . In tests on wood, in addition to the preservative effect, the effect of the evaporation factor, which is of great importance with oily preservatives, is determined with (more or less) accuracy."

In 1951 Finholt⁴⁶ revives the use of emulsion of the creosote fractions in the old Petri dish or flask method, like history repeating itself. The reader can be excused if he senses a degree of confusion.

Creosote Losses

Baechler's 1949 paper⁵ on the toxicity of oils before and after aging would be a fitting introduction to this section. The first task is to condense into simple statements or tables some of the available data on creosote losses from treated wood.

Curtin³² cites Bond on the latter's experimental determination in 1910-11 of creosote losses from thoroughly air-seasoned red oak and maple railroad ties with approximately the same moisture content. Bond reports that:

1. Full cell treated red oak in 200 days between November 1910, and June 1911, lost 19.0 per cent; and
2. Similar empty cell treated ties lost 52.7 per cent; and
3. Full cell treated maple ties in 105 days from March to June 1911, lost 13.4 per cent; and
4. Similar empty cell treated ties lost 23.0 per cent of the creosote absorbed at treatment.

The losses were determined by weight. He shows that the losses were

greater for the lighter treatments, and that in effect the losses were relatively greater for the empty cell than for the full cell treatments. His conclusions appear to hold good through all the subsequent cited data that admit of such comparisons.

Bateman's early, 1912, work⁶ on oils extracted from two old piles that had been in service about thirty years indicated some considerable loss — in one case more than 35 per cent in the above water section — and relatively lower losses for water line and below water line sections. This confirmed a generally accepted common opinion. Losses below water line were apparently confined to the fraction distilling below 225°C in the case of the pile which he considers to have been treated with a pure coal-tar creosote. He calls *light* oils those distilling below 205°C. With the exception of the above water section of one pile all the samples of treated wood still contained about 17 pounds per cubic foot. He states:

"The creosote in the pile which was perfectly preserved contained originally at least 40 per cent of naphthalene fractions, a large portion of which remained in the wood. The creosote in the pile which was less perfectly preserved contained little or no naphthalene."

Service records on such oils resemble records from the Washington-Norfolk line and from the specimens examined by Alleman¹ in that the data have little if any bearing — other than historical — on the creosote use problems of today.

Schmitz et al^{102, 103} report a loss of 25.9 per cent after five years service in track in red oak ties treated with a 60/40 creosote-coal tar solution, compared to a loss of 17.3 per cent after three years service. The losses, determined by extraction, varied inversely as the boiling range, as was expected. There was very little loss in the 315–355°C fraction, and no loss is indicated for the fraction boiling above 355°.

Bateman in 1922,⁷ and again in 1936 (see Discussion and ¹⁰⁸) in connection with his explanation of the relation between the loss of the creosote fraction below 270°C and a formula for estimating the permanence or preservative life of creosote in treated wood, cites the earlier work of von Schrenk, Fulks and Kammerer, and Rhodes and Hosford on creosote losses from southern pine poles in the Washington-Norfolk and Montgomery-New Orleans lines of the American Telephone and Telegraph Company. Certain of the poles in question were installed in 1897 and removed in 1906 after about nine years service in line. The creosote used to treat these poles is reported to have had a specific gravity of 1.022–1.030 at 3°C above the melting point of the oil. The residue above 315°C was about 16 per cent; and the per cent naphthalene was "not less than 40 per cent." Using the pitch residue — the per cent

boiling above 315°C — as a base for calculation, along with the change in residue determined from extracted creosote, von Schrenk, Fulks and Kammerer estimated the creosote losses, above and below ground line, for five poles from the Washington-Norfolk line, that are shown in Table XVIII. The poles were old growth longleaf pine, with a high heartwood volume. They were heavily treated — to about 16 pounds — by a full cell process. There is a wide variation in the results shown in the table

TABLE XVIII — CREOSOTE LOSSES, BASED ON RESIDUE INCREASE, FROM SOUTHERN PINE POLES

9 years exposure; Washington-Norfolk line, American Telephone and Telegraph Company—Data of von Schrenk et al.

Pole No.	Average loss, per cent	
	Top	Butt
1425	43.3	2.7
29	42.4	16.4
10749	50.9	20.8
2931	58.0	32.8
9700	70.8	60.3
Overall average	53.1	26.6

TABLE XIX — CREOSOTE LOSSES, BY EXTRACTION

Southern pine test posts, aerial sections; 1926 and 1927 series; Gulfport test plot—BTL (Waterman) data.

	Exposure period, months					
	8	17	22	31	32	46
	Average loss, per cent					
<i>1926 Series</i>						
8 lb empty cell.....	16		31		41	54
12 lb full cell.....	9		21		25	34
"Light" oil.....	17		32		41	55
"Mixed" oil.....	10		22		28	38
<i>1927 Series</i>						
"Light" oil						
8 lb empty cell.....		27		44		
12 lb full cell.....		29		36		
"Heavy" oil						
8 lb empty cell.....		16		36		
12 lb full cell.....		18		30		

TABLE XX — CREOSOTE LOSSES FROM SOUTHERN PINE POSTS

Analyses of original creosote and creosote extracted from outer 1 inch below ground line of posts in test 3, 6 and 7 years; Gulfport test plot; 1925 series; 12 lb. full-cell treatments — BTL (Waterman) data.

	Original creosote	After 3 years†	After 6 years	After 7 years
Specific gravity	1.037 (38/15.5)	1.056 (60/60)	1.057 (60/60)	1.059 (60/60)
Distillation, water free basis, per cent, cumulative				
To 210°C.....	4.1	0.5	0.3	0.1
210-235.....	32.0*	4.0	3.0	3.4
235-270.....	44.8	22.6	22.1	26.3
270-300.....	57.2	38.3	39.0	43.0
300-315.....	73.8	45.9	47.2	50.7
315-355.....	88.9*	72.2	73.9	75.8
Residue.....	10.7†	26.4	25.2	24.1
Total.....	99.7	98.6	99.1	99.9
Sulph. residue, gm/100 ml.....	3.2	4.2	3.7	2.4
Tar acids, gm/100 ml.....	8.3	3.3	2.8	6.5
Estimated losses, per cent, based on residue increase.....		59.5	57.5	55.6

* To 360°C in original oil analysis.

† Above 360°C in original oil analysis.

‡ The values for 3, 6 and 7 years are averages of data for 2, 3 and 3 posts, respectively.

TABLE XXI — CREOSOTE LOSSES FROM SOUTHERN PINE POSTS

Analyses of original creosote and of creosote extracted from outer 1 inch zone below ground line of posts in test 4, 5 and 6 years; Gulfport test plot; 1926 series; 8 lb. empty-cell treatments — BTL (Waterman) data.

	Original creosote	After 4 yrs.	After 5 yrs.	After 6 yrs.
Specific gravity.....	1.044 (38/15.5)	1.080 (60/60)	1.084 (60/60)	1.131 (60/60)
Distillation, water free basis, per cent, cumulative				
to 210°C.....	1.3	0.3	0.9	—
210-235.....	34.2	1.7	2.9	—
235-270.....	60.1	15.7	21.1	—
270-300.....	74.1	38.4	39.7	6.6
300-315.....	79.9	49.2	50.0	15.2
315-355.....	96.3*	79.7	77.6	54.7
Residue.....	3.4†	19.5	21.6	44.1
Total.....	99.7	99.4	99.2	98.8
Sulph. res., gm/100 ml.....	1.7	0.5	1.1	—
Tar acids, gm/100 ml.....	8.2	2.4	2.7	3.9
Estimated losses, per cent, based on residue increase.....		69.2	72.2	86.4

* To 360°C in original oil analysis.

† Above 360°C in original oil analysis.

for the butt sections, but the top and butt figures, respectively, seem to bear some relation to each other.

R. E. Waterman in a Bell Telephone Laboratories' memorandum dated January 23, 1928, reported losses of creosote from poles removed from the Montgomery-New Orleans line and in a memorandum dated March 7, 1931, reported creosote losses, determined by periodic extractions, from the aerial sections of southern pine posts treated in 1926 and 1927. Companion posts are among the earliest lots reported on by Lumsden.⁷⁶ Part of Waterman's data are condensed in Table XIX. His figures confirm in general the conclusions reached by Bond,³² namely, that the losses were greater for the light than for the heavy treatments, for the empty cell than for the full cell treatments, and in addition, for the lighter oil than for the heavier oil. Such conclusions are in line with what might be expected from the physical characteristics of the creosotes and general knowledge of the distribution and dispersion of the creosotes in the various treatments.

The losses shown in Table XIX are rounded figures that apply to the whole cross section of the pole-diameter posts. Of more significance are data on creosote losses from the *outer 1 inch* of the *below ground* section of companion posts in the Gulfport plot. Tables XX and XXI show distillation figures for the original creosotes and for the extracted oils, from full cell and from empty cell posts, after varying exposure periods up to about seven years. The oils were both low residue creosotes. The indicated percent losses are based on the increase in the residue above 355°C — of which more later. The losses are greater for the empty cell treatments than for the full cell treatments. The fact that so much of the loss occurred within the first four years is extremely important in evaluation philosophy.

Tables XXII and XXIII present data for whole cross sections of two posts that had begun to decay and that were removed for assay four years after installation at Gulfport. Table XXII shows the original analysis of the creosote and the average analysis of the extracted oil from the two posts. The indicated loss in the ground line decay area, figured from the residue increase, was *61.1 per cent*. Table XXIII shows the distribution of creosote at treatment by zones — from the outside toward the heartwood line — from extracted borings, and the distribution of creosote after removal from test, based on extraction of sectors cut from whole cross section disks. The indicated losses, figured from average over-all retention at treatment and after removal were 65.1 and 55.5 per cent, or an average of *60.3 per cent*. This figure can be considered to be in agreement with the 61.1 per cent figure cited above.

TABLE XXII — CREOSOTE LOSSES FROM SOUTHERN PINE POSTS

Analyses of original creosote and of creosote extracted from posts after 4 years in test; 1936 series; Gulfport test plot — BTL (Waterman) data.

	Original creosote	Average, extracted creosote*
Specific gravity	1.055 (38/15.5°C)	1.134 (60/60)
Distillation, water free, per cent, cumulative to 210°C.....	3.4	0.2
210-235.....	17.0	0.5
235-270.....	40.4	2.1
270-300.....	53.3	7.5
300-315.....	59.6	15.0
315-355.....	81.2	51.5
Residue.....	18.3	47.0
Total.....	99.5	98.5
Sulphonation, residue, gm/100 ml.....	4.7	3.7
Tar acids, gm/100 ml.....	7.9	12.5

* Average analyses of toluene extracted oils from disks cut adjacent to decay line; posts 273 and 280. The estimated average loss, based on residue increase, is 61.1 per cent.

TABLE XXIII — CREOSOTE LOSSES FROM SOUTHERN PINE TEST POSTS

By zones, by toluene extraction; 1936 series; Gulfport test plot — BTL (Waterman) data.

Post No.	Years in Test	Retention, by extraction lb/cu ft					
		Whole cross section	Zones, inches				Remainder of treated sapwood
			Outer ¼	Next ¼	Next ½	Next 1	
Original retentions, at treatment							
273	—	3.8*	6.9	3.5	3.8	3.9	3.0
280	—	4.7*	11.2	4.1	3.3	1.5	3.5
Retention after removal							
273	4	1.69†	1.51	1.20	1.34	1.76	2.50
280	4	1.64†§	2.78	1.32	1.90	1.63	1.10

* Average analyses of boring samples.

† Average analyses of sectors cut from a disk taken 3 inches above maximum decay line.

‡ Average analyses of sectors cut from two disks taken 3 inches above and 3 inches below the maximum decay line. (See Table XXII for analyses of original and extracted creosote.)

§ Average loss, estimated from retentions at treatment and after removal:
For whole cross section..... 60.3 per cent

For outer 1 inch

Post 273..... 70.3 per cent
Post 280..... 64.0 per cent
Average..... 67.2 per cent

Calculated losses in the outer 1 inch of these two posts — assuming 8-inch diameter — are 64.0 and 70.3 per cent, respectively, or an average of 67.2 per cent; and this figure corresponds very closely with the four-year loss figure of 69.2 per cent calculated for the empty cell posts in Table XXI. The posts represented in Tables XXII and XXIII were obviously treated to retentions that were too low to be effective; but they illustrate what is likely to happen when too low retentions of highly volatile light creosotes are used in wood in contact with the ground.

Bateman⁷ discusses a laboratory experiment to determine creosote losses, over a 70-day period from pieces of round post sections, 5 inches in diameter and 2 feet long; and he extends his comparison to data derived from experiments conducted for the San Francisco Marine Piling Commission. He reports the following treatment data for the three creosotes involved:

1. 18 lb/cu ft of a creosote with 92 per cent distilling below 275°C;
2. 10 lb/cu ft of a creosote with 40 per cent distilling below 275°C; and
3. 27.5 lb/cu ft of a creosote with 42 per cent distilling below 275°C.

His comparative loss data are condensed in Table XXIV.

Waterman and Williams¹²⁵ report creosote losses based on periodic extractions of comparable lots of specimens from treated round southern pine *saplings* exposed in the Gulfport test plot. Their data are condensed

TABLE XXIV — CREOSOTE LOSSES

By weight, from round southern pine post sections*, and from pile sections—Madison (Bateman) data.

Exposure period, days	Group†		
	18 lb/cu ft	10 lb/cu ft	27.5 lb/cu ft
	Average loss, per cent		
10	16.0	7.0	1.3
20	22.5	9.5	—
30	28.0	12.0	3.7
40	32.0	13.5	—
50	36.0	15.5	—
60	39.0	17.0	—
70	42.0	18.5	—
90	—	—	7.3
222	—	—	16.6
475	—	—	24.5
510	—	—	25.3
785	—	—	38.2

* Five inch diameter posts, 2 feet long.

† See text for group and creosote data.

TABLE XXV — CREOSOTE LOSSES FROM ROUND SOUTHERN PINE SAPLINGS

By extraction; Gulfport test plot—BTL (Waterman and Williams) data.

	Years in test					
	1		2		3	
	n	Av. loss, per cent	n	Av. loss, per cent	n	Av. loss, per cent
Empty-cell treatments; 3.5-15.1 lb/cu ft						
Above ground line.....	14	35.5	10	55.8	3	55.9
Below ground line.....	14	40.4	10	58.8	3	61.6
Full-cell treatments; 14.0-38 lb/cu ft						
Above ground line.....	9	22.7	7	44.2	4	59.8
Below ground line.....	9	13.9	7	39.2	4	42.4

TABLE XXVI — ANALYSES OF CREOSOTES USED IN WEATHERING WHEEL AND OUTDOOR EXPOSURE TESTS

Southern pine sapwood blocks—Koppers (Rhodes et al.) data.

	Creosote I	Creosote II
Specific gravity at 38/15.5°C.....	1.064	1.081
Distillation, per cent, cumulative		
0-210°C.....	3.8	1.8
210-235.....	21.2	13.7
235-270.....	41.4	30.2
270-315.....	59.8	46.3
315-355.....	80.9	64.9
Residue above 355°C.....	18.7	34.9
Water.....	0.6	0.6
Specific gravity of fractions		
235-315°.....	1.039	1.037
315-355°.....	1.106	1.104

TABLE XXVII — CREOSOTE LOSSES FROM SOUTHERN PINE SAPWOOD BLOCKS

Ether extraction; weathering wheel tests—Koppers (Rhodes et al.) data.

Exposure period, weeks	Average loss,* per cent	
	Creosote I	Creosote II
0	0.0	0.0
1	33.7	28.4
3	50.1	40.3
5	57.2	45.3
9	64.7	52.3

* Treatment retention 16.7 lb/cu ft.

in Table XXV. The loss figures are averages for all the creosotes used. It will be noted that there is a large variation in the treatment retention groups and that the losses were more rapid and definitely higher for the lower retention empty-cell specimens. Preliminary estimates from data on creosoted $\frac{3}{4}$ -inch square stakes at Gulfport indicate losses of about the same order of magnitude, with the trend in the direction of relatively higher figures than those for the round saplings. This is in line with expectation because of the use of the toluene diluent and the practice of controlling the treatments to secure lower than threshold retentions in both full-cell and empty-cell treatments.

Creosote Losses from Treated Blocks

E. O. Rhodes and his colleagues have published two excellent papers^{50, 89} that are most significant in a discussion of creosote losses from treated wood blocks. They used two creosotes, I and II, the analyses of which are shown in Table XXVI. Southern pine sapwood blocks measuring 0.5 x 0.5 x 3.0 inches, with the long axis in the direction of the grain, were treated and exposed on a weathering wheel in the laboratory, and also out of doors. The laboratory test specimens were treated to a retention of 16.7 pounds per cubic foot by soaking the blocks, heated to 105°C, in creosote at 100°C, in a sealed container. The creosote cooled during a five-hour soaking period to 40–50°C. The authors felt that the retention of about 16.5 pounds would facilitate extraction recovery of enough oil for analysis, and that "the treated portion of a tie or pole probably contains about this amount of oil." Bell Telephone Laboratories' experience has shown (Fig. 29) that the retention in the commercially treated 8-pound post averages only about 12 pounds per cubic foot in the outer $\frac{1}{4}$ inch of wood and that the retention drops off rapidly in the wood farther beneath the surface. The Rhodes blocks, therefore, must be regarded as heavily treated.

Losses of creosote were determined by ether extraction, and the change in character of the preservative was determined by distillation of the extracted oil. Corrections were made for resin extracted with the creosote. The losses of creosote from the test blocks on the weathering wheel are shown in condensed form in Table XXVII.

The authors report creosote loss from blocks treated to a 15.0-pound retention and exposed outdoors during the winter as 44.4 per cent; and similar blocks treated to a 16.5-pound retention lost 47.1 per cent in a nine-week exposure period during the summer. The results of the experiments were taken to mean that the losses were of about the same character in the outdoor winter and summer exposure tests, and that the blocks

weathered on the wheel in the same way they did outdoors. Rhodes reaffirms his conclusions about the weathering wheel tests in 1936 in a discussion of the Snell and Shipley paper¹⁰⁸ in these words:

"Our consideration of this problem convinced us, and Snell and Shipley agree with the opinion, that natural weathering produced by heat and cold, rain and wind, involves not only evaporation but water leaching and mechanical losses of whole oil by water or by bleeding. . . . To simulate these conditions, we exposed blocks of wood treated with the test creosotes to variations in temperature, to moving water and to moving air. . . . In fact, we believe that our wood-block exposure tests include most, if not all, of the factors of natural weathering. . . ."

Now to go back a bit to Curtin's 1926³² experiments, this time referring to his own weathering tests on creosoted wood. Table XXVIII is an interpretation of the results he obtained by exposing small blocks, cut from pressure treated 2 x 4 inch southern pine sapwood stakes, to natural out-of-door weathering. His procedure was extremely severe, but he was aiming at an extreme accelerated test for permanency. After treatment his 2 x 4 inch sample pieces were held in storage under cover for 2 months; and then they were cut up so that the test blocks were about 2 x $\frac{1}{8}$ x $\frac{3}{8}$ inches. These small pieces were exposed 15 feet above the ground on wooden trays for a four-month period from September, 1926, to January, 1927, and for a ten-month period from September, 1926, to July, 1927. Losses from the 2 x 4 pieces that must have occurred during their two-month storage period would increase the loss figures shown in the table. The losses are obviously greater for the lighter treatments, and for the lower residue oil.

TABLE XXVIII — CREOSOTE LOSSES FROM SOUTHERN PINE BLOCKS*

Outdoor weathering tests—Based on data by Curtin.³²

Creosote†	Trt. No.	Average retention lb/cu ft			Average loss, per cent	
		At treat- ment	After 4 mos.	After 10 mos.	After 4 mos.	After 10 mos.
1	1	23.18	19.69	17.13	15.1	26.1
1	2	17.13	11.69	10.88	31.8	36.5
1	3	19.38	6.50	5.44	37.4	47.6
2	1	26.19	18.75	15.00	28.4	42.7

* See text for description of blocks

† Creosote number 1 was a domestic oil, specific gravity 1.056 and residue above 355°C of about 26 per cent; number 2 was a British oil, specific gravity, 1.068 and residue above 355°C of about 19 per cent.

In their roof exposure, outdoor weathering tests Duncan and Richards^{35, 39} have regularly found losses of creosote, by weight, in $\frac{3}{4}$ -inch blocks treated with creosotes having residues above 355°C of 20–30 per cent, to run in the neighborhood of 45–50 per cent. All treatments were made by a full-cell vacuum process with toluene creosote solutions. The over-all exposure period consisted of three stages,³⁹ a three-week conditioning period in a constant humidity and temperature room at 30 per cent relative humidity and 80°F, a sixty-day outdoor exposure on a rack on the roof, and a three-week reconditioning to approximately constant weight in the same humidity room. The losses were figured from the original conditioned weight of the blocks, the creosote retention at treatment, and the calculated amount of creosote remaining as indicated by the weight of the weathered and reconditioned blocks. A condensed and simplified summary for two creosotes and two block shapes^{39, 41} is shown in Table XXIX.

As might be expected under identical weathering conditions, the losses were higher in the case of the $\frac{3}{8} \times \frac{3}{4} \times 1\frac{1}{2}$ inch blocks — with twice as much end grain exposed as the $\frac{3}{4}$ " drilled cubes — which presumably resulted in an accelerated longitudinal evaporation. In other words, a given percentage loss of creosote is arrived at sooner in the case of the block with greater transverse surface area. The loss increase amounts to about 7 per cent at the 8 to 10-pound retention level. This loss is about twice as great as it would be if calculated on the basis of the increased

TABLE XXIX — CREOSOTE LOSSES FROM LONGLEAF
SOUTHERN PINE SAPWOOD BLOCKS

Outdoor weathering tests—Madison (Duncan and Richards) data.

Treatment retention lb/cu ft	Creosote number		
	Coop. No. 2	5340	5340
	Creosote loss, per cent, by weight		
18*	44.0†	—†	—‡
16	45.0	—	—
14	45.9	—	—
12	47.0	—	—
10	48.0	42.5	50.0
8	49.5	44.3	51.2
6	51.3	46.5	54.6
4	54.3	49.8	60.0
2	60.0	54.5	75.0

* Retention by full-cell treatment under vacuum with toluene-creosote solutions.

† $\frac{3}{4}$ -inch cubes.

‡ $\frac{3}{8} \times \frac{3}{4} \times 1\frac{1}{2}$ inches.

surface area only of the flat blocks. The losses are interpreted in general to mean that the exposure conditions were not as severe as those in the Rhodes⁸⁹ weathering wheel tests, and in line with his outdoor tests.

The results of recent determination at Bell Telephone Laboratories of creosote losses from $\frac{3}{4}$ -inch cube blocks, by toluene extraction, are shown in Table XXX. The blocks were divided into thirds before extraction (Fig. 30) and the respective parts were further divided and then pooled in the extractor.

The blocks in each group of twenty were selected to represent the whole gradient of treatments with an average treatment retention for each group of 6.05 pounds per cubic foot. The average density of the blocks, oven-dry weight and volume basis, was the same for each group, namely 0.56. The oil used was cooperative creosote No. 11, a 50/50 blend of British vertical retort tar creosote and British coke oven tar creosote,^{5, 12, 39} diluted as usual with toluene. The exposure period outdoors was sixty days at the Chester Field Station between January 4 and March 3, 1952, plus several days exposure on a bench in a steam heated laboratory both before and after the outdoor period. The average loss of 40.3 per cent is considered to be in line with losses for the same creosote in the Madison tests, considering the factor of the winter climate at Chester.

The distribution of the residual creosote in the test blocks as shown by the averages reported may be considered to be remarkably uniform. The differences in these averages are not regarded as statistically significant. Further discussion of losses from weathered blocks will be resumed in later paragraphs.

TABLE XXX — CREOSOTE LOSSES FROM LOBLOLLY-SHORTLEAF SOUTHERN PINE SAPWOOD BLOCKS, AND LB/CU FT REMAINING

By toluene extraction; outdoor winter weathering tests — BTL (Snoke and MacAllister) data.

Lot No.	n	Average retention at treatment, lb/cu ft	lb/cu ft remaining			
			Outer* third	Middle third	Outer third	Whole block
1	20	6.05	3.62	3.43	3.93	3.66
1	20	6.05	3.49	3.49	3.68	3.55

Average lb/cu ft of creosote remaining;

in outer thirds 3.68

in middle thirds 3.56

in whole blocks 3.61

Average loss of creosote; 2.44 lb/cu ft, or 40.3 per cent.

* $\frac{3}{4}$ -inch cubes. See cutting diagram in Fig. 30.

Creosote Losses from Impregnated Filter Paper

There have been numerous criticisms of the use of creosote loss figures obtained by evaporation from open dishes in any consideration of creosote permanence.¹⁰⁸ The use of the losses reported in this section may be criticized in a similar manner, but they represent extreme acceleration and they seem to have a bearing on the interpretation of any weathering

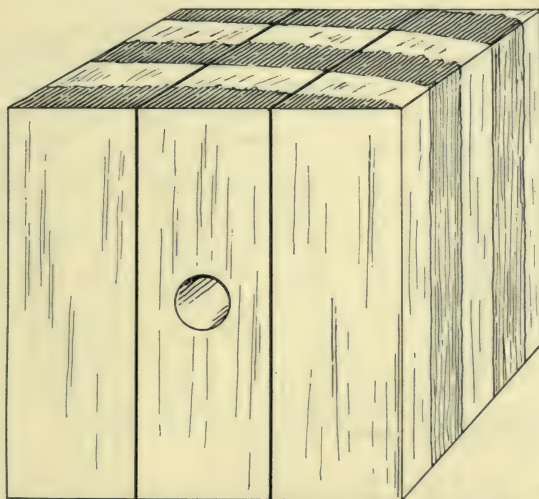


Fig. 30—Diagram of cutting plan for dividing a weathered creosoted block into three approximately equal parts for determination of residual creosote by toluene extraction.

tests in which evaporation plays the major part. Some years ago the late Heinrich T. Boving of Bell Telephone Laboratories, ran an extensive series of evaporation experiments on eight so-called Fulweiler creosotes. He used impregnated strips of filter paper, hung on quartz springs in protecting glass apparatus and exposed to a constant flow of air with no turbulence, and under constant temperature and humidity. His experiments were performed with great care. Repetitions gave excellent agreement. A condensation of his loss figures, rounded off to whole numbers, for seven- and fourteen-day exposure periods, are shown in Table XXXI. Let it be stated unequivocally at this point that it is recognized that there are important physical differences in the wood fiber combinations represented by filter paper, wood blocks, small round saplings, posts and poles; but the *quantitative* loss data seem to fall sooner or later into a similar pattern for all these test media, under the various condi-

tions, and with the assumptions made. About the difference in the *qualitative* changes that take place during creosote losses there is much, but not enough, information.

An Interpretation of Creosote Losses

Frosch,⁴⁹ in describing certain physical characteristics of the Fulweiler oils, states that they may be considered as truly viscous solutions in

TABLE XXXI — CREOSOTE LOSSES, BY WEIGHT, FROM IMPREGNATED FILTER PAPER, AND CALCULATED INCREASE IN RESIDUE
BTL (Boving) data

Fulweiler creosote No.	Original residue above 355° C, per cent	Per cent loss		Calculated residue*	
		in 7 days	in 14 days	in 7 days	in 14 days
1	43.84	31	—	63.5	—
2	40.25	32	36	59.2	62.9
3	31.11	36	—	48.6	—
4	28.52	38	—	46.0	—
5	20.85	44	—	37.2	—
6	15.49	48	—	29.8	—
7	12.21	54	—	26.5	—
8	8.49	61	68	21.8	26.5

* Assuming that all loss occurs in fraction below 355°C.

TABLE XXXII — THEORETICAL CHANGES IN CREOSOTE

Loss of volatile fractions by evaporation; amount remaining of total fraction below 355°C.

1	2	Time period 1*					Time period 2				
		3	4	5	6	7	8	9	10	11	12
Fulweiler creosote No.	Assumed treat- ment re- tention lb/cu ft	Per cent loss	Resid- ual oil lb/cu ft	Calculated per cent residue above 355°C in residual oil	More than 355°	Less than 355°	Per cent loss	Resid- ual oil lb/cu ft	Calculated per cent residue above 355°C in residual oil	More than 355°	Less than 355°
					lb/cu ft					lb/cu ft	
2	8.00	32	5.44	59.2	3.22	2.22	36	5.12	62.9	3.22	1.90
8	8.00	61	3.12	21.8	0.68	2.44	68	2.56	26.5	0.68	1.88
2	10.00	32	6.80	59.2	4.03	2.77	36	6.40	62.9	4.03	2.37
8	10.00	61	3.90	21.8	0.85	3.05	68	3.20	26.5	0.85	2.35
2	12.00	32	8.16	59.2	4.83	3.33	36	7.68	62.9	4.83	2.85
8	12.00	61	4.68	21.8	1.02	3.66	68	3.84	26.5	1.02	2.82

* Time periods 1 and 2 represent exposures that would result in the per cent losses determined from Boving's 7 and 14 day tests, respectively. See text.

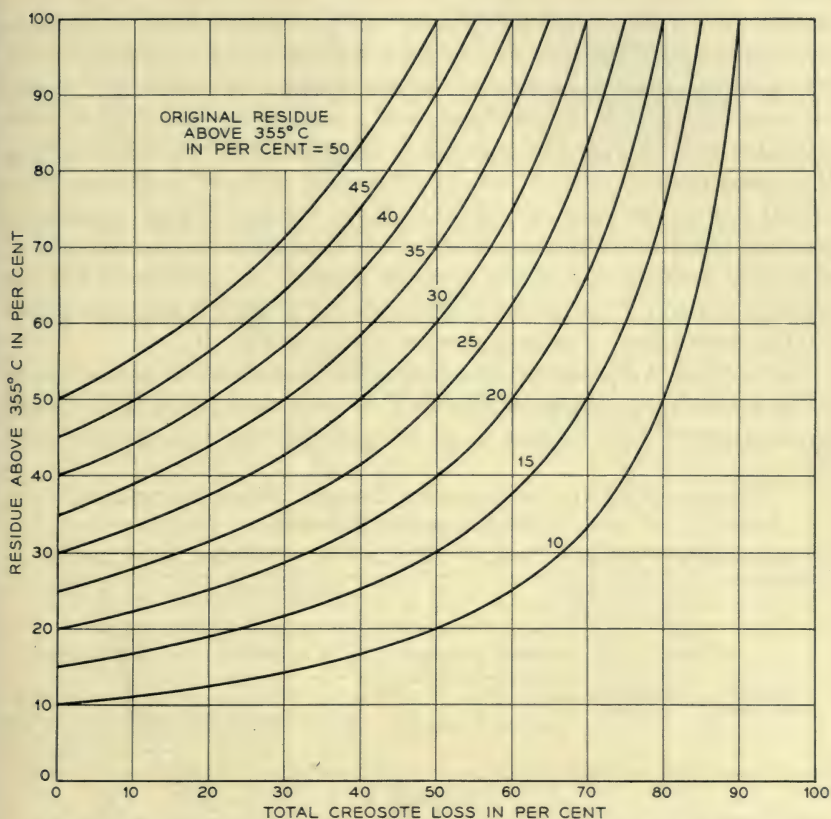


Fig. 31—Theoretical relation of the per cent increases in the residue above 355°C and the per cent losses of total creosote.

which the fraction below 355°C is the solvent and the fraction above 355°C is the solute; and that this condition does not hold for any other temperature point.

Let it be assumed that one may dodge the inferences of the results obtained by Hudson and Baechler⁵⁹ about the increase that may occur in the residue above 355°C as a result of oxidation. Here one would be in agreement with Schmitz et al.^{102 p. 270} Let it be assumed further that all loss takes place in the fraction below 355°C, and that the residue above 355°C is inert, (cf. Loseby and Krogh⁷⁸). The residue in the eight creosotes used by Boving would increase in their appropriate ratios to the figures shown for seven and fourteen days in Table XXXI. Losses for creosotes of different initial residues and consequent residue increases can then be represented by a family of curves for residue increase with

quantitative loss such as those shown in Fig. 31. Now let it be assumed that these hypotheses can be applied to treated wood, with full realization that the application may err in the direction of oversimplification. Bateman and Cislak accept the general principles of the 355°C division point between volatile and nonvolatile creosote constituents in debating the theoretical aspects of creosote losses.¹⁰⁸ Rhodes⁸⁹ indicates some actual loss in the fraction above 355°C in his test blocks, possibly in part the result of oil displacement in the water phase of his tests, rather than any increase that might occur as a result of oxidation.⁵⁹ For the purpose at hand in this paper it is convenient to use the relations shown in the percent loss — residue increase curves in Fig. 31.

Table XXXII represents a theoretical approach to what would happen to the gross characteristics of the oils if, in some time period X, the losses from creosoted wood treated to 8-, 10- and 12-pound retentions became

TABLE XXXIII — CREOSOTE LOSSES* FROM SOUTHERN PINE SAPWOOD BLOCKS†

Calculations of residual fractions below 355°C; weathering wheel tests.‡

1	2	3	4	5	6	7	8
Exposure period, weeks	Creosote loss, per cent	Creosote remaining		By extraction		By calculation§	
		per cent	lb/cu ft	Residue above 355°C, per cent	Residual fraction be- low 355°C lb/cu ft	Residue above 355°C, per cent	Residual fraction be- low 355°C lb/cu ft
Creosote I							
0	0.0	100.0	16.70	19.0	13.53	19.0	13.53
1	33.7	66.3	11.07	25.2	8.28	28.7	7.89
3	50.1	49.9	8.33	34.6	5.45	38.1	5.16
5	57.2	42.8	7.15	40.2	4.28	44.4	3.97
9	64.7	35.3	5.90	48.7	3.03	53.8	2.73
12	66.5	33.5	5.59	56.3	2.44	56.7	2.42
Creosote II							
0	0.0	100.0	16.70	33.6	11.09	33.6	11.09
1	28.4	71.6	11.96	43.7	6.73	46.9	6.34
3	40.3	59.7	9.97	51.4	4.85	56.3	4.36
5	45.3	54.7	9.14	57.3	3.90	61.4	3.53
9	52.3	47.7	7.97	60.8	3.12	70.4	2.36
12¶	55.0	45.0	7.50	65.0	2.53	74.7	1.89

* Losses based on ether extraction.

† Blocks $\frac{1}{2}$ x $\frac{1}{2}$ x 3 inches.

‡ See Bibliography, References 50 and 89.

§ Assuming that all loss occurs in the fraction below 355°C.

|| Retention by soaking in undiluted creosote.

The 12-week figures were calculated from extrapolations of the loss curves.

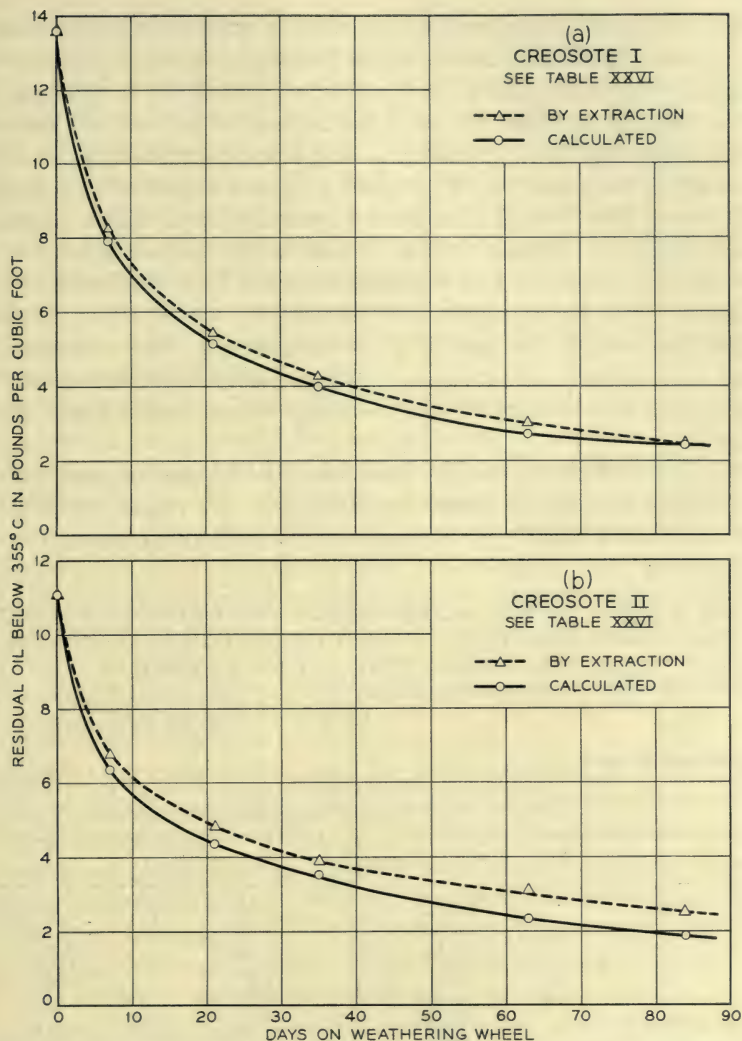


Fig. 32—Creosote losses; lb/cu ft of the fraction below 355°C remaining in the test blocks, by calculation and by ether extraction, based on weathering wheel test data. See text. (a) Creosote I, 19.0 per cent residue above 355°C. (b) Creosote II, 33.6 per cent residue above 355°C.

as great as those determined by the Boving evaporation experiments. The columns have been numbered to facilitate reference. Attention is directed to Columns 6 and 7, and to columns 11 and 12. It will be noted that at the end of Time Period 1 the calculated amounts of material boiling below 355°C in creosotes 2 and 8 are approaching the same magnitude in the respective 8-, 10- and 12-pound treatment groups; and by the end of Time Period 2 (columns 11 and 12), the calculated amounts in each respective treatment group are practically the same for these two creosotes that originally had residues above 355°C of 40.25 and 8.49 respectively. Actually, of course, the time periods would not be the same for the respective 8-, 10- and 12-pound treatments, but — on the basis of loss data already presented — would probably be relatively shorter for the lower retention group and relatively longer for the higher retention group.

Table XXXIII illustrates two methods of arriving at estimates of the per cent and amount of creosote remaining in the respective fractions above and below 355°C, by extraction data and by calculation of the

TABLE XXXIV — CALCULATED AMOUNT OF CREOSOTE FRACTION
BELOW 355°C REMAINING AFTER VARIOUS EXPOSURE PERIODS
UNDER WEATHERING WHEEL TEST CONDITIONS

Exposure period, weeks	Treatment retentions—lb/cu ft		
	16.7	10.0	8.0
	Residual fraction below 355°C lb/cu ft		
Creosote I			
0	13.53	8.10	6.48
1	7.89	4.96	3.46
3	5.16	3.44	<u>2.16</u>
5	3.97	<u>2.57</u>	1.59
9	2.73	1.82	.99
12	<u>2.42*</u>	1.46	.84
Creosote II			
0	11.09	6.64	6.11
1	6.34	4.03	<u>2.72</u>
3	4.36	2.91	1.85
5	3.53	<u>2.35</u>	1.37
9	<u>2.36</u>	1.87	.81
12	1.89	1.58	.59

* See text for significance of horizontal lines, page 485.

theoretical residue increase related to the given loss data. The loss data are those of Rhodes et al (loc. cit.) (see Table XXVII) with extrapolation to a twelve-week period; column 2. The same holds for the residue data in column 5, from which the residual pounds per cubic foot below 355°C of Column 6 were calculated, by applying the respective percent residues to the pounds per cubic foot of total creosote remaining in the blocks, column 4. The figures in column 8 were calculated in the same way by using the theoretical percent residue figures in column 7. The data in columns 7 and 8 are represented by the curves in Fig. 32, for Creosotes I and II. The curves approach each other closely enough for the purposes of the present interpretation, particularly in the case of the 19 per cent residue oil.

Now, supposing that Rhodes et al had used two lower retentions, say 10 pounds and 8 pounds in their experiments in addition to the 16.7 pounds; and assuming that their losses would be of about the same magnitudes for the 16.7 and 10-pound treatments and slightly higher for the 8-pound treatment, one would come out with theoretical calculations like those shown in Table XXXIV. The significance of the magnitudes of these amounts below 355°C will be suggested in the following paragraphs.

The Gross Characteristics of the Residual Creosotes in Soil-Block Tests of Weathered Blocks

Madison test data for average losses of creosotes by weight, from treated blocks, during the periods of weathering and reconditioning are rounded off and plotted in Fig. 33. The usual procedure was to bring the weathered blocks into the laboratory and place them under constant temperature and humidity of 80°F and 30 per cent until the blocks came to approximate weight equilibrium. Examination of the figure indicates very definitely (1) heavier losses in the lower retentions respectively for all of the creosotes, and (2) lower losses for the higher residue oils than for the lower residue oils. These conclusions follow naturally from a consideration of the distillation ranges of the oils and the treatment conditions, confirming the inferences from other loss data already discussed.

These losses of creosote from 8- to 10-pound groups of weathered treated wood blocks seem to be in the neighborhood of 40-50 per cent by weight or by extraction. The aim of the weathering techniques now being developed at Bell Telephone Laboratories is to reach some definable end point in the protective life of the preservative that will reflect the end point or failure point of test plot specimens or poles in line.

Other things being equal, one can then consider and plan for treatment retentions sufficiently high to assure a protective residual of a given preservative for a long, economical service life of the treated plant units. This service life is not necessarily an indefinitely long life, nor is it the maximum physical life that might be obtained by using relatively larger quantities of preservatives than might be consistent with cleanliness requirements. Implicit is the idea of avoiding at all times heavy maintenance and replacement costs on account of decay, but particularly in the early life of the line. To accomplish these ends it is necessary to know how much preservative to use.

The following interpretation of the results of soil-block tests is confined to weathered block experiments, for the practical reasons previously cited. The interpretation is not offered as something entirely new; but it looks like a good working hypothesis, and it seems to help in explaining what may be happening in laboratory soil-block tests of creosote.

TABLE XXXV — SUMMARY AND INTERPRETATION OF SOIL-BLOCK TESTS

Weathered, creosoted southern pine sapwood blocks; creosote losses; amounts and gross characteristics of residual oils at threshold retentions for *Lentinus lepidus*.

1	2	3	4	5	6	7	8	9	10	11
Item	Creosote No.*	Specific gravity 38/ 15.5°C	Residue above 355°C per cent	Thresh- hold lb/cu ft	Per cent loss	Creosote loss lb/cu ft	Residual creosote lb/cu ft	Calculated residue above 355°C	Residual creosote	
									>355°C lb/cu ft	<355°C lb/cu ft
1	1	1.065	18.5	9.8	53.1	5.2	4.6	39.4	1.81	2.79
2	7	1.077	20.5	9.0	47.8	4.3	4.7	39.3	1.85	2.85
3	2	1.081	30.6	10.2	47.1	4.8	5.4	57.8	3.12	2.28
4	6	1.093	34.2	9.0	37.8	3.4	5.6	55.0	3.08	2.52
5	3	1.108	50.4	12.2	30.3	3.7	8.5	72.3	6.15	2.35
6	8	1.115	53.2	9.4	25.5	2.4	7.0	71.4	5.00	2.00
7	9a		21.2	5.7	40.4	2.3	3.4	35.6	1.21	2.19
8	9	1.001	20.0	5.8	43.1	2.5	3.3	35.1	1.16	2.14
9	10a		14.4	6.7	50.9	3.4	3.3	29.4	0.97	2.23
10	10	1.068	15.2	6.9	52.2	3.6	3.3	31.8	1.05	2.25
11	11	1.038	18.0	6.5	47.7	3.1	3.4	34.4	1.17	2.23
12	M1	1.107	41.9	8.0	33.8	2.7	5.3	63.3	3.35	1.95
13	M2	1.070	18.1	8.3	50.6	4.2	4.1	36.6	1.51	2.59
14	BTL 5340	1.088	20.9	7.5	46.6	3.5	4.0	39.1	1.57	2.43

* Creosotes 1, 2, 3, 6, 7, 8, 9, 10 and 11 are those in use in the Cooperative Creosote Tests (see Bibliography, References 12 and 39). Oils 9a and 10a are samples from the same lots as numbers 9 and 10. (See Bibliography, Reference 36.) For oils M1 and M2 see Bibliography, References 37 and 38. Creosote 5340 is shown in Table II.

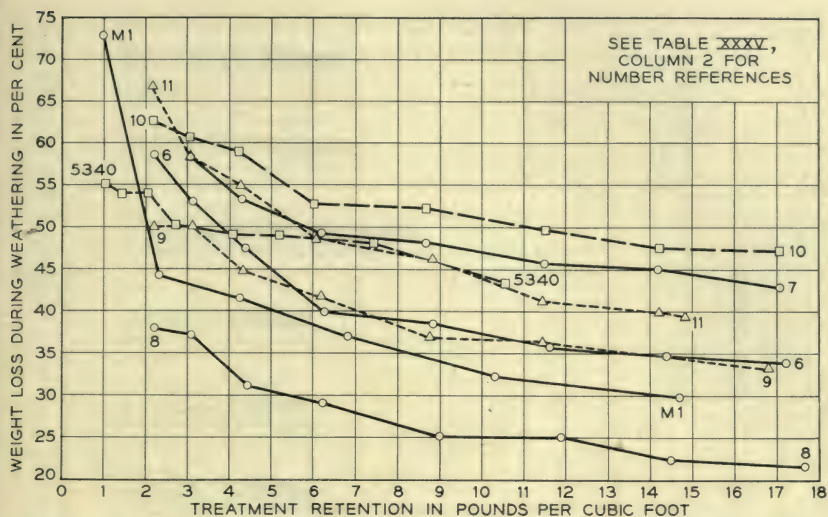


Fig. 33—Creosote losses in weathered, $\frac{3}{4}$ -inch cube, southern pine sapwood test blocks; relation of per cent loss of total creosote to original residue above 355°C and retention at treatment, lb/cu ft. All treatments were made with toluene creosote solutions. The total elapsed time from treatment to final reconditioned weight was about 105 days, including 60 days outdoor exposure on the Laboratory roof at Madison, Wis. See Table XXXV for number references.

Table XXXV is a condensed set of data on fourteen lots of weathered, creosoted $\frac{3}{4}$ -inch cube blocks. All of the tests were run at the Forest Products Laboratory at Madison, Wisconsin, in the Division of Forest Pathology, under the direct supervision of the same investigator, Dr. Catherine G. Duncan.^{36, 37, 38, 39, 41} The test fungus was *Lentinus lepideus*, Mad. 534. Ten of the tests have been run in cooperation with Bell Telephone Laboratories, and four have been run more or less concurrently with other cooperators. The technique for handling the weathered blocks has been essentially the same, and it has been rigidly controlled, except for the vagaries of the weather itself, at all essential points.

The data for the creosotes (Cols. 3 and 4), for the thresholds (Col. 5), and for the amount of residual oil in the blocks at the time they were placed in test (Col. 8), and the per cent and amount lost (Cols. 6 and 7) are all taken from the published reports or from manuscripts either ready³⁶ or in preparation for publication.⁴¹ The writer has calculated the residues in the residual oils, and the respective amounts remaining above and below 355°C (Cols. 9, 10 and 11) in pounds per cubic foot, on the assumption that all the loss occurred in the fractions boiling below 355°C . Particular attention is directed to the figures in Col. 11 — the calculated amounts remaining of the fractions boiling below 355°C . In terms of

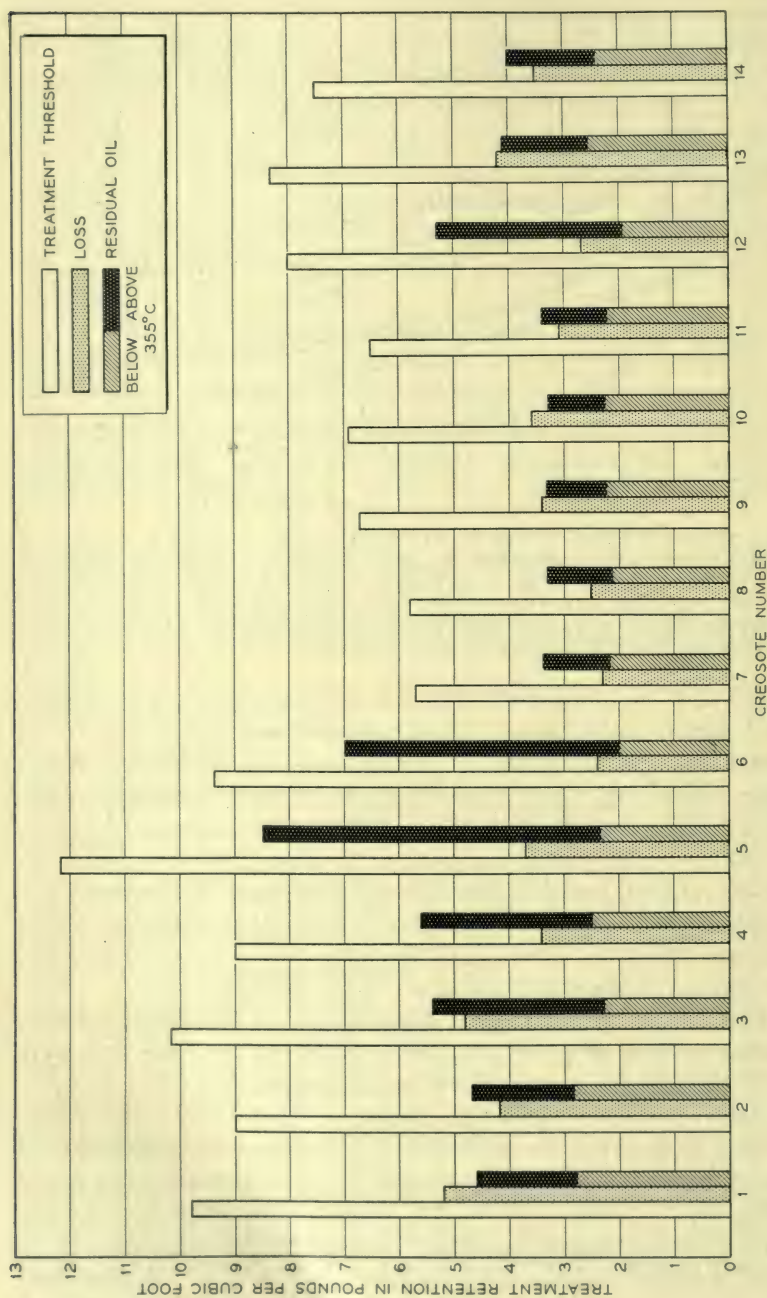


Fig. 34—Soil-block tests of fourteen creosotes against *Lentinus lepideus*; weathered southern pine sapwood blocks; comparison of treatment thresholds, losses during weathering and reconditioning, and gross characteristics of residual creosotes in terms of calculated amounts remaining of fractions above and below 355°C; based on Madison data. See Table XXXV.

pounds per cubic foot the over-all picture of the relative threshold amounts of creosote at the time of treatment, the oil lost, and the calculated proportional parts of the residual oil — after weathering — above and below 355°C, are shown graphically in Fig. 34. All of the data are in terms of pounds per cubic foot. If one bears in mind that the thresholds (Col. 5, Table XXXV) as given by the Madison investigators were located by a combination of visual observation of the blocks and extrapolation of straight lines through the weight loss data one may conclude that *in all of these tests the results were essentially the same for all the*

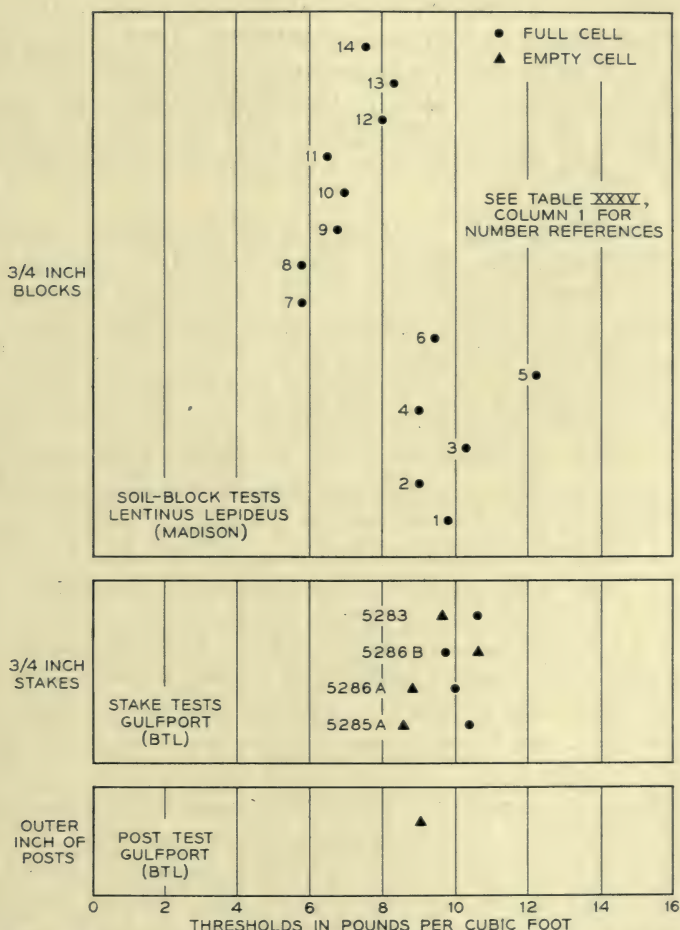


Fig. 35—Relative values of creosote thresholds by soil-block tests, $\frac{3}{4}$ -inch stake tests, and post tests.

creosotes with respect to the indicated threshold amounts distilling below 355°C. Statistically, the figures in Table XXXV Col. 11, are not significantly different. Slight changes in estimating the thresholds would conceivably bring them all to approximately the same level. In any given set of experiments the level would also vary with the duration (and types) of a different weathering cycle.

It will be recognized that no attempt has been made to separate or define the gross components of the fraction remaining below 355°C. This calls for more study, and for the development of refined methods of extraction and assay by weight and by distillation. Also, no attempt has been made to interpret the value or significance of the residue above 355°C, either because of its possible retardation of the evaporation of lower fractions, or because of some potential mechanical blocking effect. The whole interpretation is based on the simple division of the creosotes into two parts, the part distilling below 355°C and the part distilling above 355°C. Refinement will depend upon better future experimental evidence.

As an illustration of the application of the hypotheses discussed in the preceding paragraphs, one may reexamine the data from the weathering wheel experiments.⁸⁹ Rhodes, Roche and Gillander used one creosote retention only in their weathering wheel experiments, namely 16.7 pounds per cubic foot. In commenting on their work C. S. Reeve^{87 p. 78-79} noted Rhodes' emphasis on "the fact that toxicity without permanence is just as worthless as permanence without toxicity". Reeve and his colleagues carried out somewhat similar weathering experiments using slabs of wood about $\frac{1}{8}$ inch thick, exposing the treated pieces to somewhat lower temperatures than those in the Rhodes' experiments and "following a procedure with circulated air, and heat, and water". The plan of the experiments called for the conduct of "weathering cycles with reduced increments of various oils in order to get down finally to a percent of impregnation right at the end of a weathering cycle which would actually yield a rotting specimen of *Lentinus lepideus*". The work had not progressed far enough to accomplish this end, but Reeve says "The results . . . are in very close corroboration of what Mr. Rhodes has found. In other words, our loss curves with different oils running from relatively low residues to relatively high residues, have been almost parallel, I believe, with the loss curves which he has shown . . ."

Rhodes et al used essentially an agar-block method for testing their weathered blocks against *Lentinus lepideus*, Mad. 534, the same strain as that used in the Madison tests. The residual creosotes in the blocks treated with Oil I and Oil II (Table XXVI) are calculated to have been

5.90 and 7.97 pounds per cubic foot at the end of the nine-week weathering cycle (Table XXXIII); and by extraction these residuals contained 3.03 and 3.12 pounds per cubic foot of the fractions distilling below 355°C. On the basis of residue change alone these amounts are calculated at 2.73 and 2.36 pounds respectively. Rhodes states^{89 p. 76} "In no case was a weathered specimen attacked by the fungus". Paraphrasing his next sentence, this proved that both Creosote I and Creosote II *at a treatment retention of 16.7 pounds per cubic foot* "were affording adequate protection at the end of nine weeks, equivalent to many years of actual service". Was there any reason to expect such specimens to decay?

It is easier to attempt an answer to that question now than it was in 1934. Duncan⁴¹ shows a treatment threshold retention for "conditioned", i.e., unweathered, blocks of 1.6 pounds per cubic foot by agar-block tests, which is in close agreement with an average of 1.56 calculated from recent European tests reported by Schulze, Theden and Starfinger.⁵⁴⁽¹⁾ Of more importance for the question at hand is Duncan's weathered agar-block creosote threshold, given as 5.0 pounds per cubic foot. The loss in creosote at this 5 pound level was 57 per cent, which left 2.2 pounds per cubic foot of residual creosote in the blocks. The residue is calculated to have risen, as a result of weathering losses of the lower fractions, from an original 20.9 per cent (Table II) to 48.5 per cent; and on the basis of this figure the 2.2 pounds of residual oil consisted of 1.07 pounds per cubic foot of the fraction above 355°C and 1.13 pounds of the fraction below 355°C.

The Rhodes' nine-week weathered blocks still contained roughly two to three times this amount below 355°C. One may conclude that the nine weeks weathered blocks should not have shown decay under the culture conditions; and certainly none of the more briefly weathered blocks should have shown evidence of attack. If original retentions of 10 and 8 pounds of creosote had been used in the experiments the test fungus might have attacked nine and twelve weeks weathered 8-pound blocks treated with either Creosote I or Creosote II (Table XXXIV).

Using the results of the Madison tests shown in Table XXXV and Fig. 34 as indices of what might have happened if Rhodes, Roche and Gillander had used 16.7, 10.0 and 8.0 pounds at treatment and if — instead of employing an approximate agar-block technique — they had run their evaluation tests by the soil-block method, one would have expected decay to show up as indicated by the horizontal lines in the data columns in Table XXXIV. In other words, in the 16.7 pound treatments *Lentinus lepideus* would probably have attacked the blocks if they

had been weathered twelve weeks; and the 10.0 blocks would have been attacked at the end of five weeks weathering; while the 8-pound blocks would have been attacked after three weeks and about two weeks in the case of Creosote I and II, respectively. Under all these assumptions the treatment threshold for creosote in these block tests would probably have been set at above 8 pounds, and possibly around 9 pounds or more per cubic foot, in 1934; which would agree very well with the evidence obtained from soil-block tests, $\frac{3}{4}$ -inch stake tests, and test posts that has been presented in this paper.

Older records show very substantial quantities of the fraction below 355°C remaining in well treated wood after long service. Alleman's 1907 paper¹ is a classic. Writing of the increasing use of creosoted wood he states: "Recent reports . . . have clearly shown that, while proper treatment gives remarkably good results, much of this timber was not properly treated and has not lasted as it should". All of which in his opinion "... makes it imperative that we should know, as completely as possible, just what constitutes efficient creosote treatment. The different sorts of oils are believed to have different preservative values when injected into timber, but there is, unfortunately, a lack of uniformity in opinion".

Alleman chose, as the best method for finding some of the answers, an extraction of oils from treated timber that had given good service. As solvents he used absolute alcohol and subsequently anhydrous benzene. He fractionated the extracted creosotes with a view to determining the character of the oils, deciding to make his cuts so as to collect the distillate as follows: I to 170°C; II 170–205°C; III 205–245°C, which he regarded as the naphthalene fraction; IV 245–270°C; V 270–320°C; VI 320–420°C; and VII the residue above 420°C.

The wood from which he extracted the creosotes consisted of ties, mostly British, piles from England and the United States, paving blocks, and a section of creosoted wood duct removed in perfect condition after fourteen years service in Bell Telephone plant in Philadelphia. The English piles had been in service forty-three years, the other old samples all averaged a little over twenty years. Alleman's extractions showed — after all these years in use — that there were on the average over 9 pounds per cubic foot of oil remaining in the ties and English piles; nearly 9 pounds remaining in the conduit; and about 16 pounds remaining in the American piles and the paving blocks.

The writer has calculated the residue above 355°C in these extracted oils to have varied from about 23 to about 42 per cent; and the pounds per cubic foot of oil distilling below 355°C remaining in the treated wood

ran from a low of about 5.5 pounds to about 11.0 pounds. The wood had apparently remained sound. Alleman cites the difficulties of arriving at precise judgments but concludes "that 10 pounds of creosote per cubic foot is ample for railroad ties, and that piles require from 10 to 20 pounds" according to location.

Alleman's discussion of the relative amounts of *light* and *heavy* oil that might be desirable are not applicable to present day oils and commercial conditions. His perplexities remain — in almost identical form or in modernized version — and his extraction results are a long way from those reported by Lumsden,⁷⁶ and those cited elsewhere in this paper. Breazzano was beginning the application of biological tests in Italy²⁰ with a view to correlating chemical and fungicidal characteristics of preservatives. This type of endeavor was later to be pressed vigorously by Bateman,⁸ whose work has already been mentioned.

If the interpretation offered is supported by additional experiments already under way the Madison data in Table XXXV and Figs. 34 and 35 will be recognized as representing one of the most consistent series of laboratory tests for the evaluation of creosotes that has ever been run.

The Evaluation of Greensalt

The satisfactory performance of posts and poles treated with greensalt^{74, 79} has been reported in Lumsden's paper.⁷⁶ So far the very satisfactory results at the Gulfport test plot are accurate indices of what has been found by examination of poles in line. The only data on greensalt treated $\frac{3}{4}$ -inch stake tests reported in this paper are shown in Table XII and Figs. 19 and 23. Summaries of additional stake data are now in preparation for publication.

The indications are that the threshold for greensalt under Gulfport conditions is 1.42–2.1 pounds per cubic foot for $\frac{3}{4}$ -inch stakes. The average life estimates for the two treatment groups — 0.57 and 1.17 pounds of dry salt per cubic foot — (Table XIII) compare most favorably with the estimates for the four creosotes in the same table. The number of greensalt specimens is large enough to warrant the conclusion that the lines used for estimating the threshold in Fig. 23, in their trend to fall off to the right soon after the sixth year of exposure, indicate that particular period as a critical one for comparisons and interpretations. Commercially treated southern pine poles meeting the standard specification requirements for retention — 1 pound of dry salt per cubic foot — have about 2.0 pounds of dry salt in the outer inch. The agreement between the stake and post tests seems good.

In soil-block tests^{70, 95} *Poria incrassata*, *Poria monticola* and *Lenzites trabea* have been resistant to greensalt K, whereas *Lentinus lepideus* is very susceptible. The writer interprets the results of these three types of evaluation procedure, by soil-block, by 3/4-inch stakes, and by posts or poles to mean that:

1. The conditions for decay, as far as *Lenzites trabea* are concerned, are much more favorable in the soil-block culture bottle than they are in the above ground part of a pole under normal outdoor exposure conditions;

2. The incidence of attack or infection at the ground line by *Lenzites trabea*, *Poria incrassata* and *Poria monticola* at Gulfport is relatively rare; and that

3. The success of greensalt K in southern pine poles may be attributed in large part to the susceptibility of the ubiquitous *Lentinus lepideus* to the combination of salts in the greensalt preservative.

Incidentally one may cite the following example of confirming results in tests of another salt preservative. Harrow's⁵¹ experiments with soil-block tests resulted in locating a threshold for zinc chloride on unweathered blocks at 0.28 pounds per cubic foot (at treatment) for *Lenzites trabea*, Madison 617 and 0.53 for *Poria vaporaria*. Richards and Addoms⁹⁵ found approximately 0.25 for Madison 617, and approximately 0.50 for *Poria monticola*, Madison 698. These two *Porias* are possibly the same species. The similarity in the thresholds appears to be the definite result of following the same technique, rather than a haphazard coincidence.

The Evaluation of Pentachlorophenol

The highly toxic properties of pentachlorophenol have been established by exhaustive Petri dish agar toximetric tests.⁵³ A 5 per cent solution of penta in a light petroleum solvent is the preservative of reference in the recommended standard test for evaluating oil-soluble wood preservatives of the National Wood Manufacturers Association.¹⁰⁹ This test, as pointed out previously, is an agar-block test. Duncan reports⁴¹ that threshold determinations based on soil-block tests of a 5 per cent solution of penta in petroleum (cf. Tables V and VI) against *Lenzites trabea*, a critical fungus for this preservative, have not varied more than ± 0.2 pounds from 4.8 pounds per cubic foot in 7 series of weathered block experiments over a five-year period. Recent Bell Telephone Laboratories' soil-block tests confirm this result, with the same fungus and the same petroleum carrier.

These results are confirmed at the Laboratories' Gulfport test plot

(unpublished data). In two separate series — 1937 and 1938 — $\frac{3}{4}$ -inch stakes were treated at retentions slightly below the threshold cited by Duncan⁴¹ with 5 per cent solutions of pentachlorophenol in light petroleum (gas oil) and with two coal tar creosotes. The performance after six and seven years was approximately the same for both the penta solutions and the creosotes. However, completely favorable results on test posts have been reported for Gulfport⁷⁶ and Saucier, Miss., test plots. Early tests on 2 x 4 inch stakes are now being critically examined, and more tests are in progress.¹⁸ Penta treated posts are installed in the Saucier plot, where they are under periodic observation and comparison along with posts treated with the cooperative creosotes.¹² All of these experiments will greatly facilitate correlation of the results of different test methods.

As far as pole line tests are concerned one can only echo the report⁷⁶ that up to this time not one of the tens of thousands of poles in line that were treated with either straight penta-petroleum or with mixtures of penta-petroleum and creosote have been reported as failing because of decay.

Swedish Creosote Evaluation Tests

Rennerfelt and Starkenberg⁸⁸ report that of fourteen stakes measuring 1.5 x 1.5 x 100 cm, that were cut from the middle and inner sapwood of creosoted Scotch pine poles, none are sound after ten years in the test plot (May, 1950). The stakes were rated as 3 with slight decay, 8 with medium decay, and 3 with severe decay. Apparently these results cannot be correlated with definite treatment retentions.

On the other hand, the same authors state that stakes measuring 2 x 5 x 50 cm, treated to an average retention of 5.55 pounds per cubic foot of creosote (undiluted) have all decayed in a 4.5-year exposure period in greenhouse decay chamber tests. Additional experiments have been started, presumably with stakes at higher retentions, "in order to determine whether it is possible to correlate results from such decay chamber experiments with the results obtained in field and service tests."

In another series of experiments Rennerfelt and Starkenberg find after seven years (as of May, 1950) that creosoted stakes measuring 2 x 5 x 50 cm are showing different degrees of resistance to wood-destroying fungi in their four different test plots. The difference in behavior in different test plots — which is more or less to be expected — holds true for salt as well as creosote treatments. Creosote and Bolidens (zinc-chromium-arsenic) are the better performing preservatives, with creosote in

the lead. However, there have been a total of three failures in the creosoted stakes treated with average retentions of 3.6 (two stakes) and 5.6 (one stake) pounds of creosote. Stakes treated to an 8.6 pound retention are showing slight to medium decay, and in two plots slight decay has been found on a total of three stakes treated to a 12.1-pound retention. The stakes are all treated without the addition of any diluent, such as toluene, to the creosote. The reader, bearing in mind the differences in the site conditions, can make interesting comparisons between the small stake test results obtained in Sweden and in the Gulfport test plot.

Of further interest, however, is the fact that in the Swedish tests, round posts treated with average retentions of 5.37 and 5.80 pounds per cubic foot are all rated as sound after seven years exposure. One can assume that at treatment the outer annual rings of such posts contained 8 pounds or more of creosote and this amount has been sufficient to protect the posts in the Swedish climate. Rennerfelt has stated personally to the writer that one would have to proceed with caution in Sweden in the direction of increasing the creosote retention for poles, on account of public reaction against bleeding. His test results — the only ones of their kind available from Europe to the writer's knowledge — seem to be in line with Bell Telephone Laboratories findings. They would be more interesting if he had used soil-block tests for correlation.

Shortening the Bioassay Test

Besides speeding up the weathering period by the use of a weathering wheel, or by the method of alternating water and controlled heat cycles now being developed at Bell Telephone Laboratories, there are two other avenues of approach to shortening the bioassay test. One is the use of thin wood veneer test units in place of wood blocks, in the methods of impregnation and exposure to fungus action proposed by Breazzano^{28, 72} and Hopkins and Coldwell.⁵⁵ Breazzano claims a maximum of accuracy because of uniform distribution in thin pieces of wood, 0.6–0.7 mm, of the preservative to be tested and because the fungus attack and passage *through* the thin strip gives a quick visual indication of the necessary protective threshold. He also claims advantages for his Italian method because it is not necessary to use any culture medium at all — he exposed his wood strips over water only (cf. Waterman et al¹²⁶) — and because no tedious weighing techniques and record making are required. His arguments are intriguing, but his method seems to be quite out of question for testing toxicity-permanence relations of volatile preservatives like creosote. Evaporation losses would be very rapid, close to those

reported by Cislak in discussion following the presentation of the Snell and Shipley paper,¹⁰⁸ and would approach those obtained by Boving in his impregnated filter paper experiments (see Table XXXI). Furthermore, the results of tests such as his, which involve the principle of inoculation by placing "fungus on wood," instead of "wood on fungus" as in the soil-block test, require a lot of translation to interpret their significance in practical wood preservation. Rabanus⁸⁶ brought this matter out into the debate very clearly 20 years ago. Liese et al⁷² answer Breazzano's objections to the block test.

There have been no further reports on the procedures followed by Hopkins and Coldwell. Their methods are subject to some of the same criticisms that have been mentioned in the preceding paragraphs, particularly if one were to consider such techniques in a search for a way to speed up the culture tests. However, whether one agrees with them or not, their introduction of the idea of applying strength tests leads directly to a discussion of strength losses, as against weight losses, as criteria for establishing preservative thresholds.

Toughness or Impact Tests for Determining Preservative Effectiveness

Trendelenburg¹¹⁵ published in 1940 his scheme for testing the strength of treated blocks that had been exposed to fungus attack. He was aiming at a technique that would shorten the time period of fungus tests on wood, but he was also looking for some other criterion than weight loss as a measure of fungus attack. He used the relative impact strength values of matched sound and decaying test specimens as indices of the degree of decay. Boards were carefully quarter-sawed first into pieces of double specimen width plus saw kerf tangentially, and of double specimen length. From these blanks four test specimens were cut that measured 8.5 x 8.5 x 120.0 mm. The pairs were considered to be matched laterally and vertically since every effort was made to cut them from the same annual rings. One piece from each pair was exposed to fungus attack and the other served as a control. The fungus cultures were made in Kolle flasks on malt agar. Specimens were placed radial side down directly on the growing fungus surface. In the pendulum testing machine the impact load was always applied to the upper radial face, so that the lower more or less infected radial face represented the tension side of the specimen in the breaking test.

Trendelenburg showed that the per cent change in strength caused by the test fungus in the early stages of decay was much more pronounced than the concomitant changes in weight or density. He called

attention to the fact that the German Standard³³ for testing wood preservatives contains a stipulation that a weight loss of less than 5 per cent shall not be considered significant unless there is visual evidence of actual wood destruction by the test fungus. He presents data to show that a loss of 50 per cent of the relative original impact strength in spruce and fir occurred after about fifteen days in test, and that the weight loss for the test pieces was *about 3 per cent only*. Impact strength seems to be affected much more than the bending or compression strength. He was confident that his method would not only shorten the time of the bioassay test but also give more reliable and more significant results than those based on weight loss alone.

After Trendelenburg's death his ideas have been further tested and developed by von Pechmann and Schaile.¹¹⁹ In addition to trying out the suitability of the strength test procedure, they have explored the changes in the wood structure with the microscope, and, as decay progressed, they have determined the gross relation between weight loss and solubility in sodium hydroxide.

They present as an example comparable data for the German Standard agar-block test and a test run by Trendelenburg's method, using pine wood (presumably *Pinus sylvestris*) and the test fungus *Coniophora cerebella*, against a proprietary preservative. The absorption of the preservative was essentially the same in each test; but not enough preservative was used to permit determination of the threshold. The main results were that in the impact strength test procedure in fifty days the strength reduction was 66.4 per cent and the weight loss 12.7 per cent; whereas in the standard agar-block procedure in four months the weight loss was only 2.0 per cent. Von Pechmann and Schaile feel that it is possible to save $2\frac{1}{2}$ to 3 months time by using the strength test technique, and that with proper attention to detail the results will be more definite and just as reliable as those obtained in the longer period required for the standard agar-block tests.

The precise Trendelenburg technique has not been tried out in this country in any comparative tests on wood preservatives but toughness test data on small specimens of wood and veneer, sound, fungus stained and decayed, treated and untreated, have been accumulating at the Forest Products Laboratory at Madison, Wis. The fact that strength loss begins earlier and may increase more rapidly than weight loss or than change in specific gravity in natural infections in the heartwood was shown by the writer²⁷ and his colleagues shortly after the end of World War I. Confirming data were secured in later tests.^{28, 98} Scheffer⁹⁷ showed the same results, on a more definite basis, by growing *Polystictus versicolor*

on red gum sapwood in large test tubes and testing matched specimens for various strength properties as decay progressed.

The Trendelenburg technique has possibilities, but it will be some time before one can say whether it is practicable to take full advantage of it in developing supplemental bioassay tests. The one outstanding difficulty in the way of extensive use of strength tests on small specimens lies in the procurement, in the very variable southern pine, for example, of a requisite quantity of straight grained quarter sawed wood for the manufacture of the *matched blocks*. Small scale check tests are practicable, according to the writer's experience. The cost of personal supervision and manufacture of any large number of specimens would appear in advance to be exorbitant. Still, strength loss as a result of attack in treated wood is important, and Trendelenburg's ideas may win more proponents, if only as a supplemental procedure, after the soil-block technique has become more firmly established and appreciated.

OTHER ACCELERATED BIOASSAY TESTS

There are a number of items that must be mentioned before bringing this long paper toward its conclusion. There are, for example, other types of outdoor exposure tests on wood and of laboratory block tests than those cited. Two types only will be used as illustrations of the efforts that are being made to evaluate preservatives by other than the traditional service test, namely: Verrall's^{120, 121} *above ground* outdoor testing procedure, and the experiments of Tipppo et al¹¹⁴ with large block tests devised to determine effective concentrations of preservatives for prevention or control of decay in wooden ships.

Since the soil-block test is essentially a laboratory simulation of controlled ground line conditions, there is a need for some other type of test that will approximate the above ground conditions to which treated wood may be exposed. Verrall treats pieces of dressed nominal 2 x 4 inch southern pine sapwood and exposes them to the varying wet and dry, hot and cold weather conditions at Saucier, Miss. One of two pieces has a 45° end cut. This end cut is toe-nailed to the side of the other piece, which is then nailed upright on a supporting treated or untreated rail support, with the V up. This permits maximum hazard as far as catching water is concerned. His results are furnishing valuable information about pentachlorophenol, copper naphthenate and organic mercury compounds, for example. His techniques are applicable to other preservative problems, and other investigators are using his scheme in Canada for general studies,^{105, 106} and at Ann Arbor, Mich., for testing the amount

of preservative needed in the upper, or above ground section, of thin sapwood poles.

The work of Tippo and his associates represents one phase of an extensive set of experiments in which large ($6 \times 5 \times \frac{7}{8}$ inch) and small ($3 \times 5 \times \frac{7}{8}$ inch) specimens are made up to simulate a butt-block assembly, and exposed to the attack of certain critical fungi by adding to the block assembly another inoculum block ($3 \times 5 \times \frac{7}{8}$ inch) that has been thoroughly infected. The assembled units are kept in a warm and practically saturated atmosphere until the reaction of the fungus to the different preservatives can be determined. This work is being expanded in view of the importance of minimizing decay in wooden ships.

The point to be made here is that Verrall's tests and Tippo's tests should be evaluated carefully before the service test program is broadened extensively.

OTHER OBSERVATIONS

Some of the results of Suolahti's interesting studies¹¹¹ on the influence of wood at a distance on the intensity and direction of growth of fungus filaments (mycelium) have been confirmed by preliminary experiments at Bell Laboratories. Small sterilized southern pine sapwood blocks enclosed in either test tube or Petri dish cultures exert a positive pull on the filaments that is effective over a distance of several centimeters. The growth of the mycelium is more luxuriant, and the filaments are definitely drawn in the direction of the wood. Without attempting any interpretation of the significance of this phenomenon one may be permitted to point out that such studies strongly support the very great desirability — if not the necessity — of using wood in any studies directed toward evaluation of wood preservatives.

In view of the nearly forty years of prior work both here and in Europe, in which it was definitely established that certain higher fungi were the principal causes of decay in wood, it is hard to see why Weiss¹²⁸ spent so much time and such careful work on trying to test wood preservatives by using bacteria as his bioassay agents. Following the presentation of his paper before the Society of Chemical Industries in 1911 some of Weiss's critics pointed out that his methods were unrealistic as far as oil preservatives were concerned, one of his commentators suggesting that the proper approach to the problem of preservative evaluation was *to test treated and untreated wood*, (unsterilized and sterilized) under conditions favorable for fungus growth.

Tamura in 1931¹¹² used an assembly of two pieces of treated wood

molding between which he inserted a properly sized piece of untreated wood, the whole being exposed over the surface of an agar culture of the test fungus. He did not attempt to add a block of infected wood to his setup as Trippo did; but his procedure illustrates an attempt of some twenty years ago to test the protective action of preservatives in the laboratory. His statement that sterilization might drive off a significant amount of volatile preservative from freshly treated blocks, but that the sterilizing process would probably have very little effect on the preservative residual in weathered blocks anticipated similar views expressed in the present paper.

For testing initial toxicities one can still use the Petri dish or agar flask method; but it is about as unrealistic as Weiss's procedure as far as tests of toxicity and permanence of wood preservatives are concerned. The results can be presented for their academic interest, and the investigator can keep safely aloof from the perilous practical problems of wood preservation unless he attempts to translate his data into terms of *permanence and preservative value*. Then his practical colleagues as well as his technical friends point out to him, truly with a vengeance, the error and unrealistic character of his efforts.

It may be, as Rabanus⁸⁶ has suggested, that closely similar results can be obtained by the agar and by the agar-block method in the case of certain definite toxic chemicals, particularly water soluble ones. If so, the Petri dish or agar-flask method could be used with such preservatives, and the results of the tests could be applied in practice, *after* such agreement between agar and block tests was firmly proven and established.

It is completely unrealistic to attempt to arrive at significant values for the volatile fractions of creosote by confining them within a tightly closed culture dish. If such materials are really transient their toxic function can operate only during the early life of the treated wood. In a whole creosote, for example, the relatively higher toxicity of the volatile low boiling fractions supplies an important initial power to the preservative, which power is evidently diminished as the volatiles leave the treated wood. The degree of change in toxicity, as measured by the agar method, in new and aged or weathered creosotes, as shown by the works of Snell and Shipley,¹⁰⁸ Schmitz et al,^{102, 103} Baechler⁵ and others is distinctly realistic as an index of how an oil may be altered by time and weather — at least with respect to a measure of its toxic properties. Such changes have great practical significance when minimum quantities of a preservative are employed, either for reasons of economy or in order to insure cleanliness in the treated product. The trouble is that the results of the Petri dish or agar-flask method do not indicate directly *how much*

preservative to use with a view to providing the necessary residual and effective preservative.

The agar-block tests are somewhat better in this respect. While additional proof is necessary the results of such tests may be good indices of retention requirements for certain water-borne preservatives. The comparison of agar-block and soil-block tests made by Warner and Krause¹²² is incomplete, and therefore somewhat unsatisfactory, particularly in view of the title of their article. They do not follow through, and they repeat some of the inferential objections raised by others as to the effect of soil differences and methods of interpretation. The comparison by Finholt et al^{46, 47} is also inconclusive chiefly because of the nature and design of their experiments. So also, but to a lesser degree, are the comparisons that one may draw from the first two papers on the soil-block and agar-block tests from Madison,^{95, 35} which after all, were in the nature of preliminary or reconnaissance studies, introducing the first trials of an outdoor weathering technique, and developing the necessary steps in the broader and more comprehensive plans followed later.

Duncan⁴¹ has now brought out a full scale comprehensive study of the agar-block and soil-block techniques, which, however, deals with oil type preservatives only. She shows definitely — as was indicated in the earlier Madison work — that the test fungi are more aggressive under the more natural and more realistic environment of the soil-block cultures. Definite evidence of the very important better control of moisture content in the soil-block tests is presented. The soil-block thresholds for the different preservatives are generally higher in the soil-block tests, although the *order* of effectiveness is essentially the same.

Sedziak's¹⁰⁶ recent paper comparing results of his tests on buried soil-blocks and results of tests by a soil-block technique approximate to that used at Madison and at Bell Laboratories is not convincing with respect to the implied superiority of the buried block method. His paper covers work begun after the early soil-block was started at Bell Telephone Laboratories, but before the extensive experiments at Madison were initiated. Satisfactory comparison of the work of the Madison and Ottawa laboratories is difficult because Sedziak has used a different set of test organisms, including the European *Coniophora cerebella* for example, and *Lenzites sepiaria*, which is apparently not a satisfactory discriminating organism for creosote and pentachlorophenol. He has omitted the very critical *Lentinus lepideus* that has been employed at Madison since 1944. While he interprets threshold retentions for penta and for copper naphthenate that are close to those obtained at Madison, the steps in his gradient retentions leave one wishing that the Madison and Ottawa

plans could have been more closely harmonized before the most recent Ottawa work was started, at least to the point of some tests with the same procedures and same test fungi. However, nothing in Sedziak's results negates the general conclusions reached at Madison and at Bell Laboratories about the value of the present soil-block technique for the testing of creosote and other oil type preservatives.

CONCLUSIONS

1. In the course of this paper evidence and interpretations have been presented to show that the soil-block technique incorporating a weathering procedure is a practical, rapid method of bioassay and that the results obtained from this method are in general agreement with accelerated stake and long time pole-diameter post tests on the same or similar preservatives. For example, it is shown that a creosote retention at treatment of 9-10 pounds per cubic foot is necessary to insure a satisfactory degree of preservative permanence in test blocks, in $\frac{3}{4}$ -inch stakes and in the outer 1 inch of test posts. There is no reason to believe that this minimum limit does not apply to the outer 1 inch of poles in line.

2. The good reputation of well creosoted material is reaffirmed by these findings. Moreover, they show why failures have occurred and indicate what should be done to forestall such failures.

3. Since the results of the block tests are essentially the same as the results of the much longer stake and post tests, the block test data can be used at once as a basis for the establishment of the necessary amounts of the respective preservatives distribution in the wood where they will do the most good. The possibilities of bleeding increase as the retention is increased, so the bioassay technique becomes an essential tool for closer appraisal of effective wood preserving power.

4. It is important now to recognize that the soil-block results with creosote, for example, reveal the fact that the results of the European agar-block tests are — in all cases — too low to represent indices of actual requirements in treated wood. Therefore, the results for creosote tabulated by Schulze, Theden and Starfinger^{54 (1)} have to be corrected upward by some multiplying factor; perhaps of the order of three or four, before they can be correlated with the results on blocks, stakes and posts presented in this paper. For the true scientific solution of the problems of these different techniques, perhaps an international task force may be required.

5. The interpretations presented in this paper indicate that the use of

the Laboratories' controlled weathering procedure will provide a means for determining truly effective threshold retentions for oil-type and salt-type preservatives for comparable service requirements.

6. These threshold determinations are supplying data that would have been most valuable in planning retention gradients in small stake and pole diameter post tests in test plots; and, in general, the bioassay tests explain and confirm the unequivocal results of experience.

7. Through the soil-block test a ready method is available for use in the quality control of a wood treater's current product at plants where large supplies of preservatives are received from one or more constant sources and are stored in bulk. No present method of bioassay control is sufficiently rapid to be effective or practicable on mixed samples taken at treating plants receiving preservatives at frequent intervals in small lots from varied sources.

8. The soil-block development may soon make it possible to reach approximately the ideal in which the long-time service tests of treated material in line will confirm the Laboratories' rapid test results with respect to preservative requirements. Coupled with results-type requirements (wherein the end product — not the steps of manufacture and treatment — are defined) viz., (a) retention *in the wood*, (b) penetration, and (c) cleanliness, there will then be even better assurance than at the present of the quality performance always expected under Bell System specifications.

ACKNOWLEDGEMENTS

It is hoped that the preceding pages will be accepted by the reader for what they are — condensed results of teamwork over the years — into which have been woven some individual ideas, opinions and interpretations. The writer is responsible for the literature review and for the selection of the items discussed. In one way or another various members of the Timber Products Group of Bell Telephone Laboratories have contributed to the collection and analysis of the supporting data. Thanks are due especially for help on the soil-block section to J. Leutritz, Jr., L. R. Snoke and Ruth Ann MacAllister; on the $\frac{3}{4}$ -inch stake section to J. Leutritz, Jr.; on the text post and pole line sections to G. Q. Lumsden and A. H. Hearn; on the review and editing of the manuscript to R. J. Nossaman, F. F. Farnsworth and G. Q. Lumsden; on the zealous checking and assembly of test, tables and figures to Jean E. Perry; and on the correlation of the Madison cooperative test data to Catherine G. Duncan. Dorothy Storm's untiring efforts as a secretarial task force are deeply

appreciated. Without such help from these people the paper could not have been prepared in its present substantial form and substance.

BIBLIOGRAPHY

1. Alleman, G., Quantity and Character of Creosote in Well-Preserved Timbers. Circ. 98, Forest Service, U. S. Dept. Agric., May 9, 1907.
2. Alliot, H., Méthode d'Essais des Produits Anticryptogamiques. Inst. Nat. du Bois Bull. Techn. 1, 1945.
3. Amadon, C. H., Recent Observations on the Relation between Penetration, Infection and Decay in Creosoted Southern Pine Poles in Line. Am. Wood Preservers' Assoc., Proc., **35**, pp. 187-197, 1939.
4. Baechler, R. H., Relations between the Chemical Constitution and Toxicity of Aliphatic Compounds. Am. Wood Preservers' Assoc., Proc., **43**, pp. 94-111, 1947.
5. Baechler, R. H., The Toxicity of Preservative Oils Before and After Artificial Aging. Am. Wood Preservers' Assoc., Proc., **45**, pp. 90-95, 1949.
6. Bateman, E., Quantity and Quality of Creosote Found in Two Treated Piles After Long Service. Circ. 199, Forest Service, Forest Products Laboratory Series, U. S. Dept. Agric., May 22, 1912.
7. Bateman, E., Coal-Tar and Water-Gas Tar Creosotes: Their Properties and Methods of Testing. U. S. Dept. Agric., Bul. No. 1036, pp. 57-65, Oct. 20, 1922.
8. Bateman, E., A Theory on the Mechanism of the Protection of Wood by Preservatives. Am. Wood Preservers' Assoc., Proc., 1920, p. 251; 1921, p. 506; 1922, p. 70; 1923, p. 136; 1924, p. 33; 1925, p. 22; 1927, p. 41.
9. Bavendam, W., Die pilzwidrige Wirkung der im Holzschutz benutzten Chemikalien. Mitteilungen d. Reichsinst. f. Forst und Holzwirtschaft. No. 7, 20 pp. Aug., 1948.
10. Baxter, D. V., *Pathology in Forest Practice*. 2nd Ed. XI + 601 pp. John Wiley and Sons, New York, 1952.
11. Bertleff, V., Prüfung der Fungiziden Eigenschaften Ostrauer Steinkohlenteer Imprägnieröle und ihrer Bestandteile. Chem. Zeitung, **63**, p. 438, 1939.
12. Bescher, R. H., and others, Cooperative Creosote Tests. Am. Wood Preservers' Assoc., Proc., **46**, pp. 68-79, 1950.
13. Bientfait, J. L., and T. Hof, Buitenproeven met geconserveerde palen. I ste mededeling. Circ. 8 Ser. III, Conservering en Veredeling No. 3, Central Instituut voor Materiaal Onderzoek. Delft, Holland, Dec., 1948.
14. Bland, D. E., A Study of Toxicity of Australian Vertical Retort Creosote Oils to *Lentinus Lepideus* Fr., *Polystictus Versicolor* (L) Fr., and Madison 517. Australian Council Sci. and Ind. Jl. **15**, pp. 135-146, 1942.
15. Blew, J. O., Comparison of Wood Preservatives in Stake Tests. Am. Wood Preservers' Assoc., Proc., **44**, pp. 88-119, 1948.
16. Blew, J. O., C. A. Richards, and R. H. Baechler, Evaluating Wood Preservatives. For. Prod. Res. Society, Proc., **5**, pp. 230-238, (Wood-block tests, pp. 235-236), 1951.
17. Blew, J. O., Comparison of Wood Preservatives in Mississippi Post Study (1952 Progress Report). Report No. R1757, Forest Products Laboratory, U. S. Forest Service, Madison, Wis.
18. Blew, J. O., Comparison of Wood Preservatives in Stake Tests (1952 Progress Report). Report No. D1761, Forest Products Laboratory, U. S. Forest Service, Madison, Wis.
19. Boyce, J. S., *Forest Pathology*. 550 p. McGraw-Hill Book Company, Inc., New York, 1948.
20. Breazzano, A., Metodo Biologico di Controllo dei Sistemi di Preservazione, dei Legnami Adottato dall' Istituto Sperimentale delle FF. SS. (English translation: Biological Method of Testing Systems of Wood Preservation Adopted by the State Railway Experiment Institute.) Tech. Rev. Italian Railways, **4**, No. 5, pp. 3-8, Nov., 1913. Forest Products Laboratory, Madison, Wis.

21. Breazzano, A., Metodi Normali di Prova sulla Putrescibilit  dei Legnani. (English translation: Standard Methods of Testing the Putrescibility of Wood.) Extract from Report of the Ninth Meeting of the Italian Association for the Study of Building Materials; Turin; April, 1922. Forest Products Laboratory, Madison, Wis.
22. Breazzano, A., Osservazioni sul Metodo dei Blocchetti di Legno in Usa nell'Analisi Tossimetrica delle Sostanze Conservatrici del Legno. *Revista Technica Delle Ferrovie Italiane*. **47**, No. 6, June 15, 1935. (Observations on the Wood Block Method in the Toximetric Analysis of Wood Preservatives. Excerpt from Technical Review of Italian Railways. English translation; Forest Products Laboratory.)
23. Breazzano, A., Methods for Determining the Fungicidal Power of Wood Preservatives. International Assn. for Testing Materials, Proc., London Congress, Group C—Organic Materials, Sub. group 3, pp. 484-486, April, 1937.
24. British Standard Method of Test for the Toxicity of Wood Preservatives to Fungi. British Standards Institution. British Standard No. 838, 17 pp., April, 1939.
25. Broekhuizen, S., Onderzoekingen over de Conserverende Waarde van een Aantal Houtconserveermiddelen. Rapp. Comm. Gebruikw. inh. hout. Deel II, Bijlage II, pp. 89-122, The Hague, 1937.
26. Cartwright, K. St. G., and W. P. K. Findlay, *Decay of Timber and its Prevention*. VI plus 294 p. London: His Majesty's Stationery Office, 1946. Reprinted 1948.
27. Colley, R. H., The Effect of Incipient Decay on the Mechanical Properties of Airplane Timber. (Abstract) *Phytopathology*, **11**, p. 45, 1921.
28. Colley, R. H., T. R. C. Wilson, and R. F. Luxford, The Effect of *Polyporus Schweinitzii* and *Trametes Pini* on the Shock-resistance, Compression Parallel to Grain Strength, and Specific Gravity of Sitka Spruce. Forest Products Laboratory, Project L-243-J1, Typewritten Report, 29 p., 32 figs., 4 plates. July 3, 1925.
29. Colley, R. H., and C. H. Amadon, Relation between Penetration and Decay in Creosoted Southern Pine Poles. *Bell Sys. Tech. J.*, **15**, pp. 363-379, July, 1936.
30. Colley, R. H., Some Observations on the Selection and Use of Modern Wood Preservatives. Reports, Twenty-fourth Session, Communications Section, Association of American Railroads, October, 1947, pp. 17-25.
31. Colley, R. H., Wood preservation and Timber Economy. Forest Products Institute of Canada. Papers presented at the First Annual Convention, Ottawa, Oct. 30-31, 1950.
32. Curtin, L. P., B. L. Kline, and W. Thordarson, V—Weathering Tests on Treated Wood. *Ind. and Eng. Chem.* **10**, No. 12, pp. 1340-1343, Dec., 1927.
33. DIN (Deutsche Normen) DVM 2176, Blatt 1. Pr fung von Holzschutzmitteln. Mykologische Kurzpr fung (Kl tzen Verfahren). Berlin, Aug. 1939. (New Edition DIN 52176).
34. DIN DVM 52176, Blatt 2. Pr fung von Holzschutzmitteln. Bestimmung der Auslaugbarkeit. Berlin, May 1941. (Reprinted Oct., 1948).
35. Duncan, C. G., and C. A. Richards, Methods of Evaluating Wood-Preservatives: Weathered Impregnated Wood Blocks. *Am. Wood Preservers' Assoc., Proc.*, **44**, pp. 259-264, 1948.
36. Duncan, C. G., A comparison of Two English Creosotes Produced from Coke-oven Coal Tar and Vertical-retort Coal Tar. Mss. Office Report, Division of Forest Pathology, Bur. Pl. Ind., Forest Prod. Lab., Madison, Wis., Jan. 21, 1949.
37. Duncan, C. G., and C. A. Richards, Evaluating Wood Preservatives by Soil-block Tests: 1. Effect of Carrier on Pentachlorophenol Solutions; 2. Comparison of a Coal Tar Creosote, a Petroleum Containing Pentachlorophenol or Copper Naphthenate and Mixtures of Them. *Amer. Wood Preservers Assoc., Proc.*, **46**, pp. 131-145, 1950.
38. Duncan, C. G., and C. A. Richards, Evaluating Wood Preservatives by Soil-Block Tests: 3. The Effect of Mixing a Coal Tar Creosote and a Penta-

- chlorophenol Solution with a Petroleum; a Creosote with a Coke Oven Tar or Pentachlorophenol Solution. Amer. Wood Preservers' Assoc., Proc., **47**, pp. 264-274, 1951.
39. Duncan, C. G., and C. A. Richards, Evaluating Wood Preservatives by Soil-Block Tests: 4. Creosotes. Amer. Wood Preservers' Assoc., Proc., **47**, pp. 275-287, 1951.
40. Duncan, C. G., Evaluating Wood Preservatives by Soil-Block Tests: 5. Lignite-Tar and Oil-Tar Creosotes. Amer. Wood Preservers' Assoc., Proc., **48**, 1952.
41. Duncan, C. G., Soil-Block and Agar-Block Techniques for Evaluation of Oil-Type Wood Preservatives: Creosote, Copper Naphthenate and Pentachlorophenol. Division of Forest Pathology, Bur. Pl. Ind., Forest Products Lab, Madison, Wis., Special Release No. 37, Jan., 1953.
42. Eden, Johan, och Erik Rennerfelt, Undersökningar enligt klotsmetoden av några främpregneringsmedel. (Studies on Wood Preservatives, According to the Block Method) Meddelanden från Statens Skogsforskningsinstitut, Bd. **35**, No. 10, 1946.
43. Eden, Johan, och Erik Rennerfelt, Fält och rötkammarförsök avsedda att utröna skyddsverkan hos olika främpregneringsmedel. (Field and Decay-chamber Experiments to Ascertain the Protective Effect of Various Wood Preservatives.) Meddelanden från Statens Skogsforskningsinstitut. Bd. **38**, No. 4, 1949.
44. Falek, R., Die wichtigsten reinen Holzschutzmittel, die Methoden ihre Wertzahlen, Eigenschaften und Anwendung. Hausschwammforschungen, **8**, pp. 18-20, 1927.
45. Findlay, W. P. K., A Standard Laboratory Test for Wood Preservatives. British Wood Preserving Assoc., Jl., **5**, pp. 89-93, 1935.
46. Finholt, R. W., Improved Toximetric Agar-dish Test for Evaluation of Wood Preservatives. Anal. Chem., **23**, No. 7, pp. 1039-1039, July, 1951.
47. Finholt, R. W., M. Weeks, and C. Hathaway, New Theory on Wood Preservation. Ind. and Eng. Chem., **44**, No. 1, pp. 101-105, Jan., 1952.
48. Flerov, B. C., und C. A. Popov., Methode zur Untersuchung der Wirkung von antiseptischen Mitteln auf holzerstörende Pilze. Angew. Bot. **15**, pp. 386-406, 1933.
49. Frosch, C. J., V—The Correlation of Distillation Range with Viscosity of Creosote. Physics, **6**, pp. 165-170, May, 1935.
50. Gillander, H. E., C. G. King, E. O. Rhodes, and J. N. Roche, The Weathering of Creosote. Ind. and Eng. Chem., **26**, No. 2, pp. 175-183, Feb., 1934.
51. Harrow, K. M., Toxicity of Water-Soluble Wood-Preservatives to Wood-Destroying Fungi. New Zealand Jl., Sec. B., **31**, No. 5, pp. 14-19, Mar., 1950.
52. Harrow, K. M., Note on the Soil Moisture Content Used with the Leutritz Technique for Testing Toxicity of Wood Preservatives Against Fungi. New Zealand Jl. Sci. and Tech., **4**, pp. 39-40, Jan., 1951.
53. Hatfield, I., Information on Pentachlorophenol as a Wood Preserving Chemical. Am. Wood Preservers' Assoc., Proc., **40**, pp. 47-65, 1944.
54. Holzschutzmittel Prüfung und Forschung III. Wissenschaftliche Abhandlungen der Deutschen Materialprüfungsanstalten: Berlin-Dahlem. II Folge, Heft 7, 132 p. Springer-Verlag, Berlin, Göttingen, Heidelberg, 1950.
1. Schulze, B., G. Theden u K. Starfinger, Ergebnisse einer vergleichenden Prüfung der Pilzwidrigen Wirksamkeit von Holzschutzmitteln. pp. 1-40.
2. Becker, G., Ergebnisse einer vergleichenden Prüfung der insektötenden Wirkung von Holzschutzmitteln. II Teil. pp. 40-62.
3. Becker, G., Prüfung der "Tropeneignung" von Holzschutzmitteln gegen Termiten. pp. 62-76.
4. Becker, G. u B. Schulze, Laboratoriumsprüfung von Holzschutzmitteln gegen Meerwasserschädlinge. pp. 76-83.
55. Hopkins, C. Y., and B. B. Coldwell, Surface Coatings for Rotproofing Wood. Canadian Chemistry and Process Industries. N. R. C. No. 1256, Dec., 1944.
56. Howe, P. J., Weathering and Field Tests on Treated Wood. Amer. Wood Preservers' Assoc., Proc., 1928, p. 192.

57. Hubert, E. E., A study of Laboratory Methods Used in Testing the Relative Resistance of Wood to Decay. Univ. of Idaho Bulletin, **34**, No. 15, July, 1929.
58. Hubert, E. E., *An Outline of Forest Pathology*. 543 pp. John Wiley and Sons, New York, 1931.
59. Hudson, M. S., and R. H. Baechler, The Oxidation of Creosote—Its Significance in Timber Treating Operations. Amer. Wood Preservers' Assoc., Proc., **36**, pp. 74-112, 1940.
60. Hudson, M. S., *Poria radiculosa*, a Creosote Tolerant Organism. For. Prod. Res. Soc., J., **2**, No. 2, pp. 73-74, June, 1952.
61. Humphrey, C. J., and R. M. Flemming, The Toxicity to Fungi of Various Oils and Salts, Particularly Those Used in Wood Preservation. Bul. No. 227, U. S. Dept. Agric., Aug. 23, 1915.
62. Hunt, G. M., and G. A. Garratt, *Wood Preservation*. VIII, 457 p. McGraw-Hill, New York, 1938. (In course of revision.)
63. Hunt, G. M., and T. E. Snyder, An International Termite Exposure Test—Twenty-First Progress Report. Amer. Wood Preservers' Assoc., Proc., **48**, 1952.
64. Jacquiot, C., Controle de l'Efficacité des Fongicides Utilisés pour l'Imprégnation des Bois. Étude Critique de la Technique Standard Anglaise et de la Norme Allemande DIN DVM 2176. Principes pour l'Établissement d'une Norme Française. Extr. d'Ann. l'Ecole Eaux Forêts, **8**, pp. 185-206, 1942.
65. Kaufert, F. H., A Survey of Laboratory Methods Used in the Evaluation of Wood Preservatives. Report of Committee P-6, Appendix A. Amer. Wood Preservers' Assoc., Proc., **45**, pp. 55-59, 1949.
66. Leutritz, J., Jr., The Toxic Action of Various Compounds on The Fungus *Lentinus Lepideus* Fr.). Unpublished Thesis, Columbia University, Nov., 1933.
67. Leutritz, J., Jr., Laboratory Tests of Wood Preservatives. Bell Lab. Record, **16**, No. 9, pp. 324-328, May, 1938.
68. Leutritz, J., Jr., Acceleration of Toximetric Tests of Wood Preservatives by the Use of Soil as a Medium. Phytopathology, **39**, No. 10, pp. 901-903, Oct., 1939.
69. Leutritz, J., Jr., Outdoor Tests of Wood Preservatives. Bell Lab. Record, **22**, No. 4, pp. 179-182, Dec., 1943.
70. Leutritz, J., Jr., A Wood-Soil Contact Culture Technique for Laboratory Study of Wood-Destroying Fungi, Wood Decay and Wood Preservation. Bell Sys. Tech. J., **25**, No. 1, pp. 102-135, Jan., 1946.
71. Liese, Nowak, Peters, Rabanus, Krieg, and Pflug, Toximetrische Bestimmung von Holzkonservierungsmitteln. Angew. Botanik, pp. 484-488, Nov.-Dec., 1935.
72. Liese, J. et al., Toximetrische Bestimmung von Holzkonservierungsmitteln. Angew. Chemie, **48**, Beihefte 11, 1935.
73. Loseby, P. J. A., and P. M. D. Krog, The Persistence and Termite Resistance of Creosote and Its Constituent Fractions. Jour. South African Forestry Assoc., J., No. 11, pp. 26-32, June, 1944.
74. Lumsden, G. Q., and A. H. Hearn, Greensalt Treatment of Poles. Amer. Wood Preservers' Assoc., Proc., **38**, pp. 349-361, 1942.
75. Lumsden, G. Q., Proving Grounds for Telephone Poles. Bell Lab. Record, **22**, pp. 12-14, Sept., 1943.
76. Lumsden, G. Q., A Quarter Century of Evaluation of Wood Preservatives in Poles and Posts at the Gulfport Test Plot. Amer. Wood Preservers' Assoc., Proc., **48**, 1952.
77. Lutz, M. L., Méthodes Permittant de Déterminer la Résistivité des Bois Bruts ou Immunisés Soumis à l'Attaque par les Champignons Lignicoles. Ann. l'Ecole Nat. Eaux Forêts **5**, pp. 317-327, 1935.
78. Mahlke-Troschel-Liese, Holzkonservierung (Wood Preservation), 3rd Ed. XII + 571 p. Springer-Verlag, Berlin/Göttingen/Heidelberg, 1950.
79. McMahon, W., C. M. Hill, and F. C. Koch, Greensalt—A New Preservative for Wood. Amer. Wood Preservers' Assoc., Proc., **38**, pp. 334-348, 1942.
80. Martin, S. W., Characterization of Creosote Oils. Amer. Wood Preservers' Assoc., Proc., **45**, pp. 100-130, 1949.

81. Mayfield, P. B., The Toxic Elements of High Temperature Coal Tar Creosote. Amer. Wood Preservers' Assoc., Proc., **47**, pp. 62-85, 1951.
82. Narayanamurti, D., V. Ranganathan, Ragbir Singh, T. R. Chandrasekhar, and A. Banerjee, Studies on Coal Tar Creosote as a Wood Preservative, Part II. Indian Forest Bulletin, No. 144, 1948, 7 + 43 pp., JI. of India Press, Calcutta, 1950.
83. Peters, F., W. Krieg and H. Pflug, Toximetrische Prüfung von Steinkohlen-teeröl. Chem. Zeit., **61**, pp. 275-285, 1937. (English Edition, Pub. Int. Adv. Off. Wood Pres. The Hague. 1937.)
84. Preservative Treatment of Poles. (Condensed from report by American Telephone and Telegraph Company of Aug. 3, 1931.) Am. Wood Preservers' Assoc., Proc., p. 237, 1932.
85. Preservatives Committee; Report of Committee P-6. Am. Wood Preservers' Assoc., Proc., **48**, 1952.
86. Rabanus, Ad., Die Toximetrische Prüfung von Holzkonservierungsmitteln. Angew. Bot. 13, p. 352-371, 1931. (Partial translation in English. Am. Wood Preservers' Assoc., Proc., pp. 34-43, 1933.)
87. Reeve, C. S., Comment on Creosote-Permanence Toxicity Relationships. Am. Wood Preservers' Assoc., Proc., p. 78-79, 1934.
88. Rennerfelt, Erik, och Bo Starkenberg., Träskyddskomittens fält- och röt-kammarförsök. (Field and Decay-Chamber Experiments with Wood Preservatives.) Meddelanden från Statens Skogsforskningsinstitut, Bd., **40**, No. 4, 1951.
89. Rhodes, E. O., J. N. Roche, and H. E. Gillander, Creosote Permanence-Toxicity Relationships. Am. Wood Preservers' Assoc., Proc., pp. 65-78, 1934.
90. Rhodes, E. O., History of Changes in Chemical Composition of Creosote. Am. Wood Preservers' Assoc., Proc., **47**, pp. 40-61, 1951.
91. Rhodes, F. H., and F. T. Gardner, Comparative Efficiencies of the Components of Creosote Oil as Preservatives for Timber. Ind. and Eng. Chem., **22**, No. 2, pp. 167-171, Feb., 1930.
92. Rhodes, F. H., and I. Erickson, Efficiencies of Tar Oil Components as Preservative for Timber. Ind. and Eng. Chem. **25**, pp. 989-991, Sept., 1933.
93. Richards, A. P., Cooperative Creosote Program; Preliminary Progress Report on Marine Exposure Panels. Am. Wood Preservers' Assoc., Proc., **48**, 1952.
94. Richards, C. A., Methods of Testing Relative Toxicity of Wood Preservatives. Am. Wood Preservers' Assoc., Proc., **19**, pp. 127-135, 1923.
95. Richards, C. A., and R. M. Addoms, Laboratory Methods for Evaluating Wood Preservatives: Preliminary Comparison of Agar and Soil Culture Techniques Using Impregnated Wood Blocks. Am. Wood Preservers' Assoc., Proc., **43**, pp. 41-56, 1947.
96. Richards, C. A., Laboratory Decay Resistance Tests—Soil-block Method. (In Cooperative Creosote Tests by R. H. Bescher et al.) Am. Wood Preservers' Assoc., Proc., **46**, pp. 71-76, 1950.
97. Scheffer, T. C., Progressive Effects of Polyporus Versicolor on the Physical and Chemical Properties of Red Gum Sapwood. U. S. Dept. Agric., Tech. Bul., No. 527, Sept., 1936.
98. Scheffer, T. C., T. R. C. Wilson, R. F. Luxford, and Carl Hartley, The Effect of Certain Heart Rot Fungi on the Specific Gravity and Strength of Sitka Spruce and Douglas-Fir. U. S. Dept. Agric., Tech. Bul., No. 779, 24 pp., May, 1941.
99. Schmitz, H., Laboratory Methods of Testing the Toxicity of Wood Preservatives. Ind. and Eng. Chem., Anal. Ed., **1**, No. 7, pp. 76-79, April, 1929.
100. Schmitz, H., and Others, A Suggested Toximetric Method for Wood Preservatives. Ind. and Eng. Chem., Anal. Ed., **2**, p. 361, 1930.
101. Schmitz, H., and S. J. Buckman, Toxic Action of Coal-Tar Creosote. Ind. and Eng. Chem., **24**, No. 7, pp. 772-777, 1932.
102. Schmitz, H., W. J. Buckman and H. von Schrenk, Studies of the Biological Environment in Treated Wood in Relation to Service Life. Changes in the Character and Amount of 60/40 Creosote-Coal Tar Solution and Coal Tar

- and Decay Resistance of the Wood of Red Oak Crossties after Five Years Service. *Am. Wood Preservers' Assoc., Proc.*, **37**, pp. 248-297, 1941.
103. Schmitz, H., H. von Schrenk, and A. L. Kammerer, Studies of the Biological Environment in Treated Wood in Relation to Service Life, III. *Am. Wood Preservers' Assoc., Proc.*, **41**, pp. 153-179, 1945.
 104. Schulze, B., and G. Becker, Untersuchungen über die pilzwidrige und insektentötende Wirkung von Fraktionen und Einzelstoffen des Steinkohlenteeröls. *Holzforschung.*, **2**, No. 4, pp. 95-127, 1948.
 105. Sedziak, H. P., The Wood-Block Soil Method of Accelerated Testing of Wood Preservatives. Report of Committee P-6, Appendix B. *Am. Wood Preservers' Assoc., Proc.*, **46**, pp. 55-58, 1950.
 106. Sedziak, H. P., The Evaluation of Two Modern Wood Preservatives. *For. Prod. Res. Soc., Proc.*, 1952.
 107. Snell, W. H., The Use of Wood Discs as a Substrate in Toxicity Tests of Wood Preservatives. *Am. Wood Preservers' Assoc., Proc.*, **25**, pp. 126-129, 1929.
 108. Snell, W. H., and L. B. Shipley, Creosotes—Their Toxicity, Permanence and Permanence of Toxicity. *Am. Wood Preservers' Assoc., Proc.*, **32**, pp. 32-114, 1936.
 109. Standard of N. W. M. A., Method for Testing the Preservative Property of Oil-soluble Wood-Preservatives by Using Wood Specimens Uniformly Impregnated. *Nat. Wood Mfg. Assoc.*, M-1-51, April 27, 1951.
 110. STAS 650-49, Incercarea Toxicității Substanțelor de Impregnat Contra Ciupercilor. *Comisiunea de Standardizare (Rumania)* April 1, 1950.
 111. Suolahti, Osmo, Über Eine das Wachstum von Fäulnispilzen Beschleunigende Chemischen Fernwirkung von Holz. *Statens Tekniska Forskningsanstalt.* 95 p. Helsinki, Finland, 1951.
 112. Tamura, T., New Methods of Test on the Toxicity and Preservative Value of Wood Preservatives. *Phytopathologische Zeitschrift*, **3**, No. 4, pp. 421-437, 1931.
 113. Teesdale, C. H., Volatilization of Various Fractions of Creosote After Their Injection into Wood. *Circ. 188*, Forest Service, Forest Products Laboratory Series, U. S. Dept. Agric., Oct. 17, 1911.
 114. Tippo, O., J. M. Walter, S. J. Smucker, and W. Spackman, Jr., The Effectiveness of Certain Wood Preservatives in Preventing the Spread of Decay in Wooden Ships. *Lloydia*, **10**, pp. 175-208, Sept., 1947.
 115. Trendelenburg, R., Über die Abkürzung der Zeitdauer von Pilzversuchen an Holz mit Hilfe der Schlagbiegeprüfung. *Holz als Roh- und Werkstoff*, **3**, No. 12, s. 397-407, Dec., 1940.
 116. van den Berge, J. Beoordeeling van de Waarde van Fungicide Stoffen voor Houtconserveering. 183 p. N. V. Technische Boekhandel, J. Waltman, Delft, Holland, 1934.
 117. van Groenou, H. Broese, Weatheringsproeven met Houtconserveermiddelen, (Weathering Tests with Wood Preservatives). *Materiaalkennis*, **7**, No. 10, pp. 63-65, Oct., 1940.
 118. van Groenou, H. Broese, H. W. L. Rischen and J. van den Berge, Wood Preservation During the Last 50 Years. XII + 318 p. A. W. Sijthoff, Leiden, Holland, 1951.
 119. von Pechmann, H., u. O. Schaile, Über die Änderung der dynamischen Festigkeit und der Chemischen Zusammensetzung des Holzes durch den Angriff Holzerstörender Pilze. *Forstwissenschaftliches Zentralblatt*, **69**, No. 8, s. 441-465, 1950.
 120. Verrall, A. F., Progress Report on Tests of Soak and Brush Preservative Treatments for Use on Wood off the Ground. *Southern Lumberman*, Aug. 1, 1946.
 121. Verrall, A. F., Decay Protection for Exterior Woodwork. *Southern Lumberman*, June 15, 1949.
 122. Warner, R. W., and R. L. Krause, Agar-Block and Soil-Block Methods for Testing Wood Preservatives. *Ind. and Eng. Chem.*, **43**, No. 5, pp. 1102-1107, May, 1951.
 123. Waterman, R. E., Forecasting the Behavior of Wood Preservatives. *Bell Lab. Record*, **11**, No. 3, pp. 67-72, Nov., 1932.

124. Waterman, R. E., F. C. Koch, and W. McMahon, Chemical Studies of Wood Preservation. III. Analysis of Preserved Timber. Ind. and Eng. Chem., Anal. Ed., **6**, No. 6, pp. 409-413, Nov. 15, 1934.
125. Waterman, R. E. and R. R. Williams, Chemical Studies of Wood Preservation. IV. Small Sapling Method of Evaluating Wood Preservatives. Ind. and Eng. Chem., Anal. Ed., **6**, No. 6, pp. 413-418, Nov. 15, 1934.
126. Waterman, R. E., J. Leutritz, and C. M. Hill, A Laboratory Evaluation of Wood Preservatives. Bell System Tech. J., **16**, pp. 194-211, April, 1937.
127. Waterman, R. E., J. Leutritz and C. M. Hill, Chemical Studies of Wood Preservation: The Wood-block Method of Toxicity Assay. Ind. and Eng. Chem., Anal. Ed., **10**, No. 6, pp. 306-314, June, 1938.
128. Weiss, J. M., The Antiseptic Effect of Creosote Oil and Other Oils Used for Preserving Timber. Soc. Chem. Ind. J., **30**, No. 23, Dec. 15, 1911.
129. Williams, R. R., Chemical Studies of Wood Preservation. I. The Problem and Plan of Attack. Ind. and Eng. Chem., Anal. Ed., **6**, No. 5, pp. 308-310, Sept. 15, 1934.

Abstracts of Bell System Technical Papers*

Not Published in this Journal

Principles and Applications of Converters for High-Frequency Measurements. D. A. ALSBERG¹. *I.R.E., Proc.*, **40**, pp. 1195-1203, Oct., 1952. (Monograph 2030).

The heterodyne method permits measurements over wide frequency bands with the standards operating at a fixed frequency. The accuracy of such measurements depends upon the performance of heterodyne conversion transducers or converters. Design principles are derived to maximize linearity and dynamic range and minimize zero corrections. These principles have been applied to point-to-point and sweep measurements of delay, phase, transmission, and impedance.

Ferroelectric Storage Elements for Digital Computers and Switching Systems. J. R. ANDERSON¹. *Elec. Engg.*, **71**, pp. 916-922, Oct., 1952. (Monograph 2014).

These ferroelectric storage devices, although still comparatively new, show great promise. They can store up to 2,500 bits of information per square inch in a surface only a few thousandths of an inch thick with pulses less than a microsecond long.

Transistors in Switching Circuits. A. E. ANDERSON¹. *I.R.E., Proc.*, **40**, pp. 1541-1558, Nov., 1952. Corrections to Figs. 17, 18 and 19 giving synopses published in December issue, pp. 1732 and 1733.

The general transistor properties of small size and weight, low power and voltage, and potential long life suggest extensive application of transistors to pulse- or switching-type systems of computer or computer-like nature.

It is possible to devise simple regenerative circuits which perform the normally employed functions of waveform generation, level restoration, delay, storage (registry or memory), and counting. The discussion is limited to point-contact type transistors in which the alpha or current gain is in excess of unity and to a particular feedback configuration.

Such circuits, which are of the so-called trigger type, are postulated to involve negative resistance. On this basis an analysis, which approximates the negative-resistance characteristic by three intersecting broken lines, is developed. Con-

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

¹ Bell Telephone Laboratories.

clusions which are useful to circuit and device design are reached. The analysis is deemed sufficiently accurate for first-order equilibrium calculations.

Transistors having properties specifically intended for pulse service in the circuits described have been developed. Their properties, limitations, and parameter characterizations are discussed at some length.

Mobility of Electrons in Germanium. P. P. DEBYE¹ and ESTHER M. CONWELL¹. Letter to the Editor. *Phys. Rev.*, **87**, pp. 1131-1132, Sept. 15, 1952.

The Telephone Industry in National Defense. C. A. ARMSTRONG². *Telephony*, **143**, pp. 44-46, 114, Oct. 25, 1952.

Infrared Absorption in High Purity Germanium. H. B. BRIGGS¹. Letter to the Editor. *Jl. Opt. Soc. Am.*, **42**, pp. 686-687, Sept., 1952.

New Infrared Absorption Bands in p-Type Germanium. H. B. BRIGGS¹ and R. C. FLETCHER¹. Letter to the Editor. *Phys. Rev.*, **87**, pp. 1130-1131, Sept. 15, 1952.

Automatic Switching for Nation-Wide Telephone Service. A. B. CLARK¹ and H. S. OSBORNE². *A.I.E.E., Trans. Commun. and Electronics Sect.*, **2**, pp. 245-248, Sept., 1952. (Monograph 2015).

Western Electric's Service with Standards. K. B. CLARKE³. *Standardization*, **23**, pp. 332-338, Oct., 1952.

Properties of Silicon and Germanium. ESTHER M. CONWELL¹. *I.R.E., Proc.*, **40**, pp. 1327-1337, Nov., 1952.

This article provides the latest experimental information on those fundamental properties of germanium and silicon which are of device interest, currently or potentially. Electrical properties, especially carrier density and mobility, have been treated in greatest detail. Descriptive material has been provided to the extent necessary to give physical background.

Effects of Space-Charge Layer Widening in Junction Transistors. J. M. EARLY¹. *I.R.E., Proc.*, **40**, pp. 1401-1406, Nov., 1952.

Some effects of the dependence of collector barrier (space-charge layer) thickness on collector voltage are analyzed. Transistor base thickness is shown to decrease as collector voltage is increased, resulting in an increase of the current-gain factor (α) and a decrease in the emitter potential required to maintain any

¹ Bell Telephone Laboratories.

² American Telephone and Telegraph Company.

³ Western Electric Company.

fixed emitter current. These effects are shown to lead to two new elements in the theoretical small-signal equivalent circuit. One, the collector conductance (g_c'), is proportional to emitter current and varies inversely with collector voltage. This term is the dominant component of collector conductance in high-quality junction transistors. The other element, the voltage feedback factor (μ_{ec}), is independent of emitter current, but varies inversely with collector voltage. The latter element is shown to modify the elements of the conventional equivalent tee network.

Four-Terminal p-n-p-n Transistors. J. J. EBERS¹. *I.R.E., Proc.*, **40**, pp. 1361-1364, Nov., 1952.

The equivalent circuit of a $p-n-p-n$ transistor is obtained. It is demonstrated that a $p-n-p$ transistor and an $n-p-n$ transistor can be connected so that the combination has the same equivalent circuit as the $p-n-p-n$ structure. A simplified circuit is obtained which can be used when the $p-n-p-n$ transistor is connected as a hook-collector transistor. A method of adjusting the current gain of $p-n-p-n$ transistors by external means is given as well as experimental results.

Dynamics of Transistor Negative-Resistance Circuits. B. G. FARLEY¹. *I.R.E., Proc.*, **40**, pp. 1497-1508, Nov., 1952.

A general method is presented for calculating approximately the behavior of many nonlinear circuits by dividing the region of operation into subregions, within each of which the circuit may be considered linear to a good approximation. The method is applied to a high-speed transistor switching circuit as an illustrative example.

Regenerative Amplifier for Digital Computer Applications. J. H. FELKER¹. *I.R.E., Proc.*, **40**, pp. 1584-1596, Nov., 1952.

A description of the negative-resistance properties of the point-contact transistor is presented as an introduction to the description of a regenerative amplifier. The choice of circuit parameters for the amplifier is discussed and a sample design presented. The illustrative amplifier regenerates digital information at a megacycle rate and develops pulses with rise times of less than $0.05 \mu\text{sec}$. It operates from supply voltages of $+6$ and -8 volts, with a battery drain of less than 0.05 watt. A complete set of computer building blocks has been designed around the amplifier. Their use is illustrated in two computer applications.

Evidence for Domain Structure in Anti-ferromagnetic CoO From Elasticity Measurements. M. E. FINE¹. Letter to the Editor. *Phys. Rev.*, **87**, p. 1143, Sept. 15, 1952.

Optical Position Encoder and Digit Register. H. G. FOLLINGSTAD¹, J. N. SHIVE¹ and R. E. YAEGER¹. *I.R.E., Proc.*, **40**, pp. 1573-1583, Nov., 1952.

The usefulness of transistors in systems has been given a feasibility proof through the construction and operation of a six-digit position encoder and serial-

¹ Bell Telephone Laboratories.

output digit register. This system performs the functions of photoelectric encoding, pulse regeneration, digit storage, reflected-to-natural binary translation, and digit shifting by means of circuits using transistors and other semi-conductor devices. The model occupies a volume of about $\frac{1}{4}$ cubic foot, weighs seven pounds, and consumes 16 watts of power.

Comparison of Recording Processes. J. G. FRAYNE⁴, *I.R.E., Trans.*, PGA-7, pp. 5-8, May, 1952. *S.M.P.T.E., Jl.*, **59**, pp. 313-318, Oct., 1952.

The three common forms of sound recording may be classed as mechanical (disk), photographic and magnetic. All three methods are in common use today and each is employed in a field for which it appears to be peculiarly fitted. The purpose of this article is to examine briefly the factors which determine the fidelity of each method. By fidelity we mean how true the tonal range can be reproduced, the amount and nature of harmonic distortion present, the signal-to-noise ratio possible with each method, and the amount of wow or flutter that may be expected under average conditions of reproduction for each recording process.

Transistor Shift Register and Serial Adder. J. R. HARRIS¹. *I.R.E., Proc.*, **40**, pp. 1597-1602, Nov., 1952.

A small set of basic functions, such as binary memory and elementary binary logic, can be remarkably versatile; such functions are important in switching and computing. This paper describes a piece of computing equipment which can store a pair of binary numbers and add them, producing the sum a digit at a time. The equipment is built from a basic set of functional blocks, all of which are designed around transistors. This set of building blocks consists of a binary cell, a pulse amplifier, a pulse amplifier with delay, and logic circuits. The binary cell is a flip-flop; amplifiers are monostable circuits, and logic is performed in diode gates. Some interesting special features arise from the use of transistors. These features are discussed and the designs are evaluated.

Charge Transfer and the Mobility of Rare Gas Ions. J. A. HORNBECK¹. *Jl. Phys. Chem.*, **56**, pp. 829-831, Oct., 1952.

Ion-atom collisions in the rare gases between an atomic ion and a parent gas atom, such as Ne^+ and Ne, involve quantum mechanical symmetry effects which though rigorously inseparable have been listed as (a) a force of resonance attraction, (b) a force of resonance repulsion, and (c) charge exchange. Drift velocity measurements at high fields show that this complicated interaction may be represented to a good approximation by the hard sphere model of kinetic theory in which the collision cross section is several times the viscosity cross section of the atoms themselves.

Broad Band Matching with a Directional Coupler. W. C. JAKES¹. *I.R.E., Proc.*, **40**, pp. 1216-1218, Oct., 1952. (Monograph 2033).

¹ Bell Telephone Laboratories.

⁴ Westrex Corporation.

This paper presents the results of a theoretical and experimental study of a waveguide matching technique which allows a directional coupler to be located any distance away from the discontinuity causing the original mismatch and a broad-band match to still be obtained.

Design curves are included which give the required coupling coefficient of the directional coupler and the power loss for a given initial mismatch and desired vswr reduction. Experimental confirmation of the theory is also presented.

New General-Purpose Relay for Telephone Switching Systems. A. C. KELLER¹. *Elec. Engg.*, **71**, pp. 1007-1012, Nov., 1952. (Monograph 2034).

This new general-purpose electromagnetic relay, called the AF type is a wire spring relay. With variations providing slow release or marginal characteristics, it is known as the AG and AJ relay, respectively. It provides improved performance at lower cost.

Spherical Model of a Ferromagnet. H. W. LEWIS¹ and G. H. WANNIER¹. Letter to the Editor. *Phys. Rev.*, **88**, pp. 682-683, Nov. 1, 1952.

In order to interpret the properties of the tetragonal crystal $\text{ND}_4\text{D}_2\text{PO}_4$ (deuterated ADP) a thermo-dynamic treatment has been developed which relates the observed crystal structure change and the dielectric constant change at the transition temperature to the appearance of spontaneous polarization. For an antiferroelectric crystal, the average spontaneous polarization is zero, being oppositely directed for adjacent layers, but the square of the spontaneous polarization is large. This results in quadratic strain components which cause a change in the crystal structure below the transition temperature. It is shown that the change observed is consistent with an antiferroelectric arrangement with one of the α axes being the antiferroelectric axis. The dielectric constants in all three directions suffer a large drop below the transition temperature.

Piezoelectric, Dielectric, and Elastic Properties of $\text{ND}_4\text{D}_2\text{PO}_4$ (Deuterated ADP). W. P. MASON¹ and B. T. MATTHIAS¹. *Phys. Rev.*, **88**, pp. 477-479, Nov. 1, 1952. (Monograph 2036).

Transistors in Our Civilian Economy. J. W. McRAE¹. *I.R.E., Proc.*, **40**, pp. 1285-1286, Nov., 1952.

I. R. E. Editor's Note: At relatively long intervals there appear on the technical and industrial horizons devices of such broad scope and major significance that they profoundly affect the fields of their use. One of these epochal developments is the transistor, which bids fair to take its place beside the electron tube as one of the foundation stones of future communications and electronics.

It is accordingly timely and suitable that certain of the probable future industrial uses and effects of the transistor should be here analyzed in a guest editorial by an engineer especially qualified for this task, and who is a Fellow and Director of the Institute, and a Vice President of Bell Telephone Laboratories.

¹ Bell Telephone Laboratories.

Domain Properties in BaTiO₃. W. J. MERZ¹. Letter to the Editor. *Phys. Rev.*, **88**, pp. 421-422, Oct. 15, 1952.

Notes on Methods of Transmitting the Circular Electric Wave Around Bends. E. S. MILLER¹. *I.R.E., Proc.*, **40**, pp. 1104-1113, Sept., 1952. (Monograph 2037).

The tendency for energy to be converted out of the circular electric wave in bent round pipe may be avoided by one of three general approaches: (1) by removing the degeneracy between TE₀₁ and TM₁₁, (2) by converting to a normal mode of the bent guide at both ends of the bend, and (3) by utilizing dissipation in the unwanted modes to prevent power transfer to them. All three approaches are discussed. Normal attenuation in round pipe should be effective in moderating straightness requirements. Elliptical guide and special waveguide structures may be used to negotiate intentional bends; bending radii in the range one to 1,000 feet appear acceptable at 50,000 mc for waveguides 3/8-inch to 2 inches in diameter, respectively.

Multi-Element Directional Couplers. S. E. MILLER¹ and W. W. MUMFORD¹. *I.R.E., Proc.*, **40**, pp. 1071-1078, Sept., 1952. (Monograph 2038).

It is shown that the backward wave in a directional coupler is related to the shape of the function describing the coupling between transmission lines by the Fourier transform. This facilitates the design of directional couplers for arbitrary directivities over any prescribed frequency band. Tightly coupled directional couplers are analyzed in simple terms, and it is shown that any desired loss ratio, including complete power transfer between lines, may be achieved. The theories are verified using waveguide models operating at 4,000, 24,000 and 48,000 mc, and it is indicated that the work is applicable to many types of electrical and acoustic transmission lines.

Transistor Noise in Circuit Applications. H. C. MONTGOMERY¹. *I.R.E., Proc.*, **40**, pp. 1461-1471, Nov., 1952.

Linear circuit problems involving multiple noise sources can be handled by familiar methods with the aid of certain noise spectrum functions, which are described. Several theorems of general interest dealing with noise spectra and noise correlation are derived. The noise behavior of transistors can be described by giving the spectrum functions for simple but arbitrary configurations of equivalent noise generators. From these, the noise figure can be calculated for any desired external circuit.

Transistor Noise in Circuit Applications. H. C. MONTGOMERY¹. *I.R.E., Proc.*, **40**, pp. 1314-1326, Nov., 1952.

The invention of the transistor provided a simple, apparently rugged device that could amplify — an ability which the vacuum tube had long monopolized. As with most new electron devices, however, a number of extremely practical limitations had to be overcome before the transistor could be regarded as a

¹ Bell Telephone Laboratories.

practical circuit element. In particular, the reproducibility of units was poor — units intended to be alike were not interchangeable in circuits; the reliability was poor — in an uncomfortably large fraction of units made, the characteristics changed suddenly and inexplicably; and the “designability” was poor — it was difficult to make devices to the wide range of desirable characteristics needed in modern communications functions. This paper describes the progress that has been made in reducing these limitations and extending the range of performance and usefulness of transistors in communications systems. The conclusion is drawn that for some system functions, particularly those requiring extreme miniaturization in space and power as well as reliability with respect to life and ruggedness, transistors promise important advantages.

In Search of the Missing 6 Db. W. A. MUNSON¹ and F. M. WIENER¹. *Acoustical Soc. Am., Jl.*, **24**, pp. 498–501, Sept., 1952. (Monograph 2019).

The unexplained difference in sound pressure in the ear canal which appears to exist when equally loud low frequency tones are presented alternately from an earphone and from a loudspeaker has bedeviled acousticians for many years and, unfortunately, still continues to do so. There are presented here the results of some of the measurements carried out at the Bell Telephone Laboratories which show the magnitude of the effect and various attempts at explaining it. While no satisfactory explanation has been found, it is hoped that publication of these results will stimulate interest in the problem.

Nation-Wide Numbering Plan. W. H. NUNN². *A.I.E.E., Trans., Commun. and Electronics Sect.*, **2**, pp. 257–260, Sept., 1952. *Elec. Engg.*, **71**, pp. 884–888, Oct., 1952. (Monograph 2015).

At the present time a great variation in the types of telephone numbers exists. This is because of the number of telephones in communities of different sizes. With the advent of local dialing and now nation-wide dialing, a uniform numbering system has become necessary.

How to Detect the Type of an Assignable Cause. P. S. OLMSTEAD¹. *Ind. Quality Control*, **9**, pp. 32–34, **36**, Nov., 1952.

Silicon p-n Junction Alloy Diodes. G. L. PEARSON¹ and B. SAWYER¹. *I.R.E., Proc.*, **40**, pp. 1348–1351, Nov., 1952.

A new type of *p-n* junction silicon diode has been prepared by alloying acceptor or donor impurities with *n*- or *p*-type silicon. The unique features of this diode are: (a) reverse currents as low as 10^{-10} amperes, (b) rectification ratios as high as 10^8 at 1 volt, (c) a Zener characteristic in which $d(\log I)/d(\log V)$ may be as high as 1,500 over several decades of current, (d) a stable Zener voltage

¹ Bell Telephone Laboratories.

² American Telephone and Telegraph Company.

which may be fixed in the production process at values between 3 and 1,000 volts and (e) ability to operate at ambient temperatures as high as 300°C.

Hard Rubber. H. PETERS¹. *Ind. and Eng. Chem.*, **44**, pp. 2344-2345, Oct., 1952.

As judged by the literature, the general trend during the past year on the subject of hard rubber has been toward de-emphasis of fundamental research and more emphasis on use. Plastics, through substitution, continue to make gains in the field of hard rubber. A renewed interest is again shown in the use of latex ebonite for industrial applications. The patent situation appears to be unusually active and the interest in synthetic hard rubbers continues to increase.

Application of Information Theory to Research in Experimental Phonetics. G. E. PETERSON¹. *Jl. Speech and Hearing Disorders*, **17**, pp. 175-188, June, 1952.

Principles of Zone-Melting. W. G. PFANN¹. *Jl. of Metals*, **4**, pp. 747-753, July, 1952. *A.I.M.E. Trans.*, **194**, pp. 747-753, 1952. (Monograph 2000).

In zone-melting, a small molten zone of zones traverse a long charge of alloy or impure metal. Consequences of this manner of freezing are examined with respect to solute distribution in the ingot, with particular reference to purification and to prevention of segregation. Results are expressed in terms of the number, size, and direction of travel of the zones, the initial solute distribution, and the distribution coefficient.

Nonsynchronous Time Division with Holding and with Random Sampling. J. R. PIERCE¹ and A. L. HOPPER¹. *I.R.E., Proc.*, **40**, pp. 1079-1088, Sept., 1952. (Monograph 2041).

There is a general type of system in which an indefinitely large number of transmitters can have access to any of an indefinitely large number of receivers over a medium of limited bandwidth. In these systems, signal-to-noise ratio goes down as more transmitters are used simultaneously. This paper describes a particular system which sends samples by means of coded pulse groups sent at random times. The signal-to-noise ratio is good in the absence of interference and the effect of interference is minimized by holding the previous sample if a sample is lost. An experimental system worked satisfactorily and gave close to the predicted signal-to-noise ratio. Such a system might be used to provide communication and automatic switching in rural telephony, or for other applications.

Fundamental Plans for Toll Telephone Plant. J. J. PILLIOD². *A.I.E.E., Trans., Commun. & Electronics Sect.*, **2**, pp. 248-256, Sept., 1952. (Monograph 2015).

¹ Bell Telephone Laboratories.

² American Telephone and Telegraph Company.

Organization of the Engineering Profession. D. A. QUARLES⁵. *Elec. Engg.*, **71**, pp. 963, 964, Nov., 1952.

Since the organization of the American Society of Civil Engineers 100 years ago, professional engineering has assumed a major role in American life. The goal now to be attained is the closer organization of the entire engineering profession.

We begin a New Institute Year. D. A. QUARLES⁵. *Elec. Eng.*, **71**, pp. 867-868, Oct., 1952.

In an address before the recent Pacific General Meeting in Phoenix, Mr. Quarles, President of the Institute, evaluates the evolution in A.I.E.E. organization and policy as the Institute enters a new administrative year.

Mean Free Paths of Electrons in Evaporated Metal Films. F. W. REYNOLDS¹ and G. R. STILWELL¹. Letter to the Editor. *Phys. Rev.*, **88**, pp. 418-419, Oct. 15, 1952.

Single-Sideband System for Overseas Telephony. N. F. SCHLAACK¹. *Electronics*, **25**, pp. 146-149, Nov., 1952.

Single-sideband transmitter furnishes four voice channels for overseas telephone service. Pushbutton tuning permits rapid frequency shifts and load-control circuit minimizes interchannel crosstalk and out-of-band radiation. Copper-oxide and germanium varistors replace modulator tubes.

Automatic Toll Switching Systems. F. F. SHIPLEY¹. *A.I.E.E., Trans., Commun. and Electronics Sect.*, **2**, pp. 261-269, pp. 889-897, Oct., 1952. (Monograph 2015).

The new system was designed to implement the nation-wide switching plan which integrates the telephone switching network of the entire nation into a single unit. Requiring a high order of mechanical intelligence, this system is one of the most comprehensive ever devised.

Properties of M-1740 p-n Junction Photocells. J. N. SHIVE¹. *I.R.E., Proc.*, **40**, pp. 1410-1413, Nov., 1952.

The p-n junction photocell has a sensitivity of 30 ma per lumen for light of 2,400 degrees K color temperature, corresponding to a quantum yield approximately unity in the spectral range from visible to the long wave cutoff at 1.8 microns. Dark currents of a few microamperes are observed at room temperature, with a temperature coefficient of about +10 per cent per degree C. Both dark and light currents exhibit saturation in the range from 1 to 90 volts applied. The frequency response is flat into the 100-kc region. Short-circuit noise currents are observed around 20 $\mu\mu\text{a}$ in a 1-cps band at 1,000 cps. The photocell element is encapsulated in a plastic housing $\frac{1}{4} \times \frac{3}{16} \times \frac{3}{8}$ inch in dimensions.

¹ Bell Telephone Laboratories.

⁵ Sandia Corporation.

Interpretation of e/m Values for Electrons in Crystals. W. SHOCKLEY¹. Letter to the Editor. *Phys. Rev.*, **88**, p. 953, Nov. 15, 1952.

Transistor Electronics: Imperfections, Unipolar and Analog Transistors. W. SHOCKLEY¹. *I.R.E., Proc.*, **40**, pp. 1289–1313, Nov., 1952.

The electronic mechanisms that are of chief interest in transistor electronics are discussed from the point of view of solid-state physics. The important concepts of holes, electrons, donors, acceptors, and deathnium (recombination center for holes and electrons) are treated from a unified viewpoint as imperfections in a nearly perfect crystal. The behavior of an excess electron as a negative particle moving with random thermal motion and drifting in an electric field is described in detail. A hole is similar to an electron in all regards save sign of charge. Some fundamental experiments have been performed with transistor techniques and exhibit clearly the behavior of holes and electrons. The interactions of holes, electrons, donors, acceptors, and deathnium give rise to the properties of p - n junctions, p - n junction transistors, and Zener diodes. Point-contact transistors are not understood as well from a fundamental viewpoint. A new class of *unipolar* transistors is discussed. Of these, the *analog* transistor is described in terms of analogy to a vacuum tube.

Unipolar "Field-Effect" Transistor. W. SHOCKLEY¹. *I.R.E., Proc.*, **40**, pp. 1365–1376, Nov., 1952.

The theory for a new form of transistor is presented. This transistor is of the "field-effect" type in which the conductivity of a layer of semiconductor is modulated by a transverse electric field.

Since the amplifying action involves currents carried predominantly by one kind of carrier, the name "unipolar" is proposed to distinguish these transistors from point-contact and junction types, which are "bipolar" in this sense.

Regarded as an analog for vacuum-tube triode, the unipolar field-effect transistor may have a $m\mu$ of 10 or more, high output resistance, and a frequency response higher than bipolar transistors of comparable dimensions.

Control of Frequency Response and Stability of Point-Contact Transistors. B. N. SLADE¹. *I.R.E., Proc.*, **40**, pp. 1382–1384, Nov., 1952.

The frequency response and stability of point-contact transistors are determined to a large degree by control of the point-contact spacing and germanium resistivity. Stability is particularly important in amplifiers in which the impedances of the emitter and collector circuits are very small in the frequency range in which the transistor is designed to operate. Satisfactory stability has been obtained with developmental transistors having a frequency cutoff (3-db drop in the current amplification factor, α) ranging from 10 to 30 mc. These transistors operate under approximately the same dc bias conditions used with lower-frequency transistors, and have an average power gain of approximately 20 db. By means of the methods outlined, transistors which oscillate at frequencies as high as 300 mc have been made.

¹ Bell Telephone Laboratories.

Junction Transistor. M. SPARKS¹. *Sci. Am.*, **187**, pp. 29-32, July, 1952.

It is one of two forms of the remarkable device that amplifies electricity by the flow of electrons in a crystal. An account of its underlying principles and present state of development.

Telephone Answering Services. L. R. STANG⁶. *Telephony*, **143**, pp. 53-55, 123-124, Oct. 25, 1952.

Traffic Engineering Design of Dial Telephone Exchanges. J. A. STEWART¹. *Telephony*, **143**, pp. 13-16, 41-42, Oct. 18, 1952.

Low-Drain Transistor Audio Oscillator. D. E. THOMAS¹. *I.R.E., Proc.*, **40**, pp. 1385-1395, Nov., 1952.

A nine-element transistor audio oscillator is described. This oscillator operates with relatively low drain from a single 6-volt battery. The oscillator gives reliable performance with an output uniform to approximately ± 1 db with substantially all type 1768 point-contact transistors and without any circuit element adjustment required for variation in transistor parameters from unit to unit or with transistor ambient temperature.

Transistor Amplifier — Cutoff Frequency. D. E. THOMAS¹. *I.R.E., Proc.*, **40**, pp. 1481-1483, Nov., 1952.

The effect of positive feedback through the internal base resistance of a transistor on circuit cutoff frequency is considered.

Transistor Reversible Binary Counter. R. L. TRENT¹. *I.R.E., Proc.*, **40**, pp. 1562-1572, Nov., 1952.

The feasibility of performing a fairly complex switching function using a few elementary transistor circuits is illustrated and experimentally verified. The specific function discussed is reversible binary counting. The mechanism used to achieve reversibility and the circuitry within each building block is described. Operating margins and suggestions for design improvements for systems application are given.

Effect of Electrode Spacing on the Equivalent Base Resistance of Point-Contact Transistors. L. B. VALDES¹. *I.R.E., Proc.*, **40**, pp. 1429-1434, Nov., 1952.

A theoretical expression for the equivalent base resistance r_b of point-contact transistors is derived here. This expression is shown to check experimental values reasonably well if the severity of some assumptions made for purposes of analysis is considered. Electrode spacing, germanium-slice thickness, and resistivity of the semiconductor are shown to be the properties that affect r_b primarily.

¹ Bell Telephone Laboratories.

⁶ Illinois Bell Telephone Company.

Measurement of Minority Carrier Lifetime in Germanium. L. B. VALDES¹. *I.R.E., Proc.*, **40**, pp. 1420-1423, Nov., 1952.

A method for measuring the lifetime of minority carriers in germanium is described. Basically, it consists of liberating the carriers optically on a flat face of a crystal and measuring the concentration of minority carriers as a function of distance from the point of liberation. The mathematical model is analyzed and experimental results are presented here.

Drift Velocities of Ions in Krypton and Xenon. R. N. VARNEY¹. *Phys. Rev.*, **88**, pp. 362-364, Oct. 15, 1952. (Monograph 2028).

Drift velocities and mobilities of ions of Kr and Xe in their respective parent gases have been measured over a wide range of values of E/p_0 , the ratio of electric field strength to normalized gas pressure. Two ions appear in each gas identified as Kr^+ and Kr_2^+ in Kr and Xe^+ and Xe_2 in Xe. The relation that drift velocity varies as $(E/p_0)^{1/2}$ at high E/p_0 has been found to hold for the atomic ions and has been used to determine the equivalent hard sphere cross sections at high fields. The cross sections are 157×10^{-16} cm² for Kr and 192×10^{-16} cm² for Xe. The Langevin theory of mobilities gives excellent agreement with experimental results extrapolated to zero field strength provided that, in the theory, the hard sphere cross section is taken as large for the atomic ions and very small for the molecular ions. The range of the polarization forces is such as to render them insignificant in atomic ion collisions and of primary importance in molecular ion collisions.

Junction Transistor Tetrode for High-Frequency Use. R. L. WALLACE¹, L. G. SCHIMPF¹ and E. DICKTEN¹. *I.R.E., Proc.*, **40**, pp. 1395-1400, Nov., 1952.

If a fourth electrode is added to a conventional junction transistor and biased in a suitable way, the base resistance of the transistor is reduced by a very substantial factor. This reduction in r_b permits the transistor to be used at frequencies ten times or more higher than would otherwise be possible. Tetrodes of this sort have been used in sine-wave oscillators up to a frequency of 130 mc and have produced substantial gain as tuned amplifiers at frequencies of 50 mc and higher.

Nature of Solids. G. H. WANNIER¹. *Sci. Am.*, **187**, p. 39 Dec., 1952.

The theory that explains their various properties is a comparatively recent development of physics. From it practical benefits already begin to flow.

Magnetic Double Refraction at Microwave Frequencies. M. T. WEISS¹ and A. G. Fox¹. Letter to the Editor, *Phys. Rev.*, **88**, pp. 146-147, Oct. 1, 1952.

Stress Relaxation in Plastics and Insulating Materials. E. E. WRIGHT¹. *A.S.T.M. Bull.*, **184**, pp. 47-49, Sept., 1952. (Monograph 2024).

¹ Bell Telephone Laboratories.

Organic materials are in ever-increasing use for mechanical and electrical devices where satisfactory performance is required over long periods of time and under a wide assortment of atmospheric influences. Although the dimensional stability of plastics and electrical insulating materials under no-load conditions has been established reasonably well, there is an important gap in existing knowledge with regard to a material's ability to maintain adequate counter-stresses under compressive loading.

A.S.T.M. Method of Test D 621 - 51² uses a constant load system and measures the material's resistance to gross deformation. However, this fails to simulate the usual application where the material is subjected to constant deflection (such as fastening devices, inserts, etc.) and is required to maintain adequate counter-stresses for the foreseeable life of the part. Therefore, it has been necessary to integrate data from Method D 621 with long practical experience in order to extrapolate between two dissimilar systems.

This paper describes a constant-deflection procedure for direct measurement of stress relaxation or change thereof, thereby permitting evaluation in terms of a material's ability to maintain a tight assembly under conditions simulating actual use. Apparatus for carrying out the test is described and typical data illustrating its usefulness are included.

Contributors to this Issue

CHARLES CLOS, C.E., New York University, 1927; New York Telephone Company, plant extension engineering, valuation and depreciation matters, intercompany settlements and tandem and toll fundamental plans, 1927-47. Pratt Institute, Evening School, Mathematics Instructor, 1946-49. Bell Telephone Laboratories, studies on development planning for local and toll switching systems and research in switching probability, 1947-. Member of A.I.E.E., New York Electrical Society, Mathematical Association of America, A.A.A.S., American Statistical Association, Iota Alpha, and Tau Beta Pi.

R. H. COLLEY, A.B., Dartmouth College, 1909; A.M., Harvard University, 1912; Ph.D., George Washington University, 1918; Austin Teaching Fellow in Botany, Harvard University, 1910-12; Instructor in Botany, Dartmouth College, 1909-10 and 1912-16; Pathologist, Division of Forest Pathology, Bureau of Plant Industry, U. S. Department of Agriculture, 1916-28. Bell Telephone Laboratories, 1928-52. Dr. Colley was chairman of Committee 05 — Wood Poles, of the American Standards Association for nearly twenty years. He was president of the American Wood-Preservers' Association 1943-44. During his years with the Laboratories he worked particularly on development and research problems connected with material and preservative treatment specifications for poles and other timber products used in outside plant. His more recent activities were directed toward improvement of laboratory techniques for evaluating wood preservatives, and toward the development of a coordinated plan for fundamental research on oil preservatives. He was Timber Products Engineer for the Laboratories from 1940 to 1950, and Timber Products Consultant from 1950 to 1952. His article in this issue of the JOURNAL was prepared before his retirement on May 31, 1952.

KARL K. DARROW, B.S., University of Chicago, 1911. He studied at the Universities of Paris and Berlin in 1911 and 1912, specializing in physics and mathematics; Ph.D., University of Chicago, 1917. He then joined the staff of Bell Telephone Laboratories, at that time known as the Engineering Department of Western Electric Company. Here his

work has included the study, correlation, and representation of scientific information for his colleagues, keeping them informed of current advances made by workers in fields related to their own activities. As a corollary to his work, Dr. Darrow appears from time to time before scientific and lay audiences to lecture on current topics in physics and the related sciences. He has taken an active interest in education, teaching physics during summer and other sessions at Stanford, Chicago, and Columbia Universities and at Smith College. From 1944 to 1946, he served as consultant to the Metallurgical Laboratory in Chicago. Dr. Darrow is the author of *Introduction to Contemporary Physics* (1939), *Electrical Phenomena in Gases* (1932), *Renaissance of Physics* (1936), *Atomic Energy* (1948). He is a member of the American Physical Society, which he has served as secretary since 1941, the Physical Society of London, Société Française de Physique, the American Philosophical Society, of which he was a counsellor for four years, and the International Union of Pure and Applied Physics, of which he was vice president from 1947 to 1951. In 1949 he received an honorary doctorate from the Université de Lyon.

GEORGE R. FROST, Bell Telephone Laboratories, 1943-. Mr. Frost has had over twenty-three years of service with the Bell System. Since 1943 he has been an instructor at the Laboratories' School for War Training and in the Communications Development Training Program, until recently when he became a member of the Publication Department's group responsible for displays. From 1941-43 he taught communications at Fenn College, and from 1946-47 mathematics at Pratt Institute.

WILLIAM KEISTER, B.S. in E.E., Alabama Polytechnic Institute, 1930; Bell Telephone Laboratories, 1930-32 and 1936-. As a member of the switching department, he is currently preparing text material and teaching general switching circuit theory and telephone switching systems to members of the technical staff. Co-author of *The Design of Switching Circuits*, with S. H. Washburn and A. E. Ritchie (Van Nostrand, 1951). Member of A.I.E.E., Eta Kappa Nu, Tau Beta Pi, and Phi Kappa Phi.

ALISTAIR E. RITCHIE, A.B., Dartmouth College, 1935; M.A., Dartmouth College, 1937. Bell Telephone Laboratories, 1937-. As a member of the switching development group, Mr. Ritchie tested panel and crossbar circuits and made noise studies on panel and crossbar systems until

1942. He then became an instructor in the Laboratories' School for War Training. From 1945-51, he taught switching circuit design. Since 1951, in the switching engineering group, he has been working out new techniques for measuring telephone traffic in central offices. Co-author of *The Design of Switching Circuits* with W. Keister and S. H. Washburn (Van Nostrand, 1951). Member of the A.I.E.E.

H. A. STONE, JR., B.S., Yale University, 1933. Bell Telephone Laboratories, 1936-. A member of the Transmission Development Department, Mr. Stone is in charge of a group engaged in the development of inductors and loading coils and cases. He previously assisted in the development of inductors and networks for use in military radio and telephone projects, and the design of radar pulse generators. Member of the A.I.E.E.

ROGER I. WILKINSON, B.S. in E.E., 1924, Professional Engineer (honorary), 1950, Iowa State College; Northwestern Bell Telephone Company, 1920-21; American Telephone and Telegraph Company, 1924-34; Bell Telephone Laboratories, 1934-43 and 1946-. U. S. War Department, Washington and South Pacific, 1943-45. Mr. Wilkinson has been engaged in applications of the mathematical theory of probability to telephone problems. Medal for Merit, 1946. Member of A.S.E.E.; A.S.A.; Institute of Mathematical Statistics; American Society for Quality Control; Fellow, Operations Research Society of America; Associate Member of A.I.E.E.; and Member of Eta Kappa Nu; Tau Beta Pi; Phi Kappa Phi; and Pi Mu Epsilon.

THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXII

MAY 1953

NUMBER

Solderless Wrapped Connections

Introduction

J. W. MCRAE 523

Part I — Structure and Tools

R. F. MALLINA 525

Part II — Necessary Conditions for Obtaining
a Permanent Connection

W. P. MASON AND T. F. OSMER 557

Part III — Evaluation and Performance Tests

R. H. VAN HORN 591

An Improved Circuit for the Telephone Set

A. F. BENNETT 611

Automatic Line Insulation Test Equipment for Local Crossbar Systems

R. W. BURNS AND J. W. DEHN 627

Theory of Magnetic Effects on the Noise in a Germanium Filament

HARRY SUHL 647

DC Field Distribution in a "Swept Intrinsic" Semi-Conductor Configuration

R. C. PRIM 665

Transmission Properties of Laminated Clogston Type Conductors

E. F. VAAGE 695

A Coupled Resonator Reflex Klystron

E. D. REED 715

Abstracts of Bell System Papers not Published in this Journal

767

Contributors to this Issue

775

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

S. BRACKEN, *President, Western Electric Company*

F. R. KAPPEL, *Vice President, American Telephone
and Telegraph Company*

M. J. KELLY, *President, Bell Telephone Laboratories*

EDITORIAL COMMITTEE

E. I. GREEN, *Chairman*

A. J. BUSCH

W. H. DOHERTY

G. D. EDWARDS

J. B. FISK

R. K. HONAMAN

F. R. LACK

J. W. MCRAE

W. H. NUNN

H. I. ROMNES

H. V. SCHMIDT

EDITORIAL STAFF

J. D. TEBO, *Editor*

M. E. STRIEBY, *Managing Editor*

R. L. SHEPHERD, *Production Editor*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; Alexander L. Stott, Treasurer. Subscriptions are accepted at \$3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXII

MAY 1953

NUMBER 3

Copyright, 1952, American Telephone and Telegraph Company

Solderless Wrapped Connections

Introduction

By J. W. McRAE

(Manuscript received February 9, 1953)

In the telephone plant during the course of a single year, the operation of connecting a wire to a metal terminal is carried out approximately one billion times. Many of these connections are made in the factory. Others are made during the installation of equipment and a substantial number are made in the course of normal operation of the telephone plant. Successful functioning of the plant depends on trouble-free performance of each of these connections, most of which are now soldered in accordance with long-standing practice. Recently, a new technique for joining wires to terminals has been developed which will have important technical and economic advantages in the Bell System and which should have similar advantages in other fields.

The immediate need for a new connection arose with the development of the new wire spring general purpose relay.* In this relay the terminals appear in the form of closely spaced wires and the standard methods of applying connections were not satisfactory. Since the production schedule for these new relays required something like fifty million connections a year on relay terminals alone, an intensive effort was made to devise a satisfactory method for wiring. The first result was the development of a tool which could wrap a few turns of wire around the terminals of the relay, and do this efficiently on the closely-

* Keller, A. C., A New General Purpose Relay for Telephone Switching Systems. Bell System Tech. J., **31**, pp. 1023-1067, Nov., 1952.

spaced wire terminals.[†] It was soon found that similar tools could be used to advantage on existing types of terminals, and the Western Electric Company is now making extensive use of wrapping tools. Connections made with these tools are being soldered after wrapping.

In the meantime, further development work indicated the possibility that by wrapping the wire under tension around a properly shaped terminal, the need for soldering might be eliminated. Major economies appeared possible if such a solderless wrapped connection were applied extensively in wiring communication equipment. In addition, there would be freedom from trouble due to solder splashes in equipment and there would be an appreciable reduction in the consumption of tin for use in solder.

Three papers in this issue of the JOURNAL describe the present status of the program undertaken to exploit these possibilities. The paper on design indicates that practical tools for making solderless wrapped connections can be designed, built and used; the paper on analysis describes the basis for belief that the method is fundamentally sound; and the paper on evaluation indicates that the resulting connections are satisfactory for use in the telephone plant.

Other types of tools capable of cutting, skinning and wrapping the wire in one operation are under development, and the problems presented in adapting the basic techniques to other conductors, such as aluminum wire and stranded copper wire are being studied. In fact, a whole new area of development effort has been opened up.

Thus the work reported in the following papers is of interest from two points of view. On the one hand it is a record of progress in the continual quest for less expensive and more reliable equipment. On the other hand, it is an example of the broad effects which are often the result of development aimed at a specific problem. The search for a solution to the problem of making connections to a new type terminal has led to a new approach to the whole problem of wire-terminal connections.

[†] Miloche, H. A., Mechanically Wrapped Connectors. Bell Labs. Record, **29**, pp. 307-311, July, 1951.

Solderless Wrapped Connections

PART I—STRUCTURE AND TOOLS

By R. F. MALLINA

(Manuscript received February 17, 1953)

In the search for a better way of connecting wires to apparatus terminals a new joining method has been discovered. The new method not only eliminates soldering and its hazards but also reduces cost, improves quality and conserves space. In contrast to the solder joint which depends largely on human judgement and skill, the new connection is made with a calibrated tool. A degree of uniformity has been obtained which virtually eliminates the need for product inspection. The trend toward smaller apparatus and automation may now be further intensified due to the use of this new method of making electrical connections.

INTRODUCTION

Methods of joining wires to apparatus terminals for the purpose of electrical conduction can be broadly divided into two groups: solder connections and pressure connections. There are others such as welded and brazed connections; however, they are relatively few in number. The annual production of solder connections in the Bell System is estimated to be one billion. In television and radio manufacture the number of connections made per year is in the order of ten billion. Because of the high cost of manual soldering, the pressure connection is of great importance to the communication industry. One form of pressure connection — the solderless wrapped connection — will be described in this article.

In order to determine the technical and economic value of a new type of pressure connection it is necessary to compare it with those now accepted as good connections in the communication industry. A large portion of this article will, therefore, be devoted to the analysis of pressure connections some of which have been in use since the early development of the telephone.

WHAT IS A PRESSURE CONNECTION?

The chart Fig. 1 shows six typical pressure connections classified in terms of seven requirements. In this classification the screw connection, for example, meets the following requirements: large contact area, high contact force, great mechanical stability, long life, easy to disconnect. The space, however, which the screw connection occupies is large and its cost is high. In the history of electricity it is probably the oldest and best pressure connection. In the second column of the table in Fig. 1 is the plug connection. It is small in size, easy to disconnect, but has no large contact area, no high contact force, no long life, no mechanical stability and is not low in cost.

As will be shown later the solderless wrapped connection in Column 6 of Fig. 1 is indicated as meeting all seven requirements. Its main advantage over the screw connection is that it is low in cost and small in size.

CONTACT AREA

The effective contact area relative to the cross sectional area of the wire is of great importance since it controls the resistance of the connection. It must remain uniform in size, metallically bright and not be affected by temperature changes, vibration and handling.

Contact area is not easily defined. For example two flat metal surfaces having an area of one square centimeter each and brought into contact do not necessarily have a contact area of one square centimeter. If the force holding them together is small, only the high spots

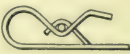
	1	2	3	4	5	6
						
REQUIREMENTS	FAHNESTOCK CLIP	PLUG	CRIMP	WIRE NUT	SCREW	SOLDERLESS WRAPPED
1. LARGE CONTACT AREA			✓	✓	✓	✓
2. HIGH CONTACT FORCE			✓	✓	✓	✓
3. LONG LIFE			✓	✓	✓	✓
4. SMALL SIZE		✓	✓			✓
5. MECHANICALLY STABLE			✓		✓	✓
6. EASILY DISCONNECTED	✓	✓		✓	✓	✓
7. LOW COST						✓

Fig. 1 — Classification of pressure connections.

of the surfaces touch and large currents passing through such a connection may develop heat and melt the metal at the high spots.

CONTACT FORCE

To make the above mentioned area of one square centimeter effective for electrical conduction it is necessary to press the two metal parts together with a force so high that essentially all particles of the area are intimately interlocked and free from insulating impurities. If the pressure is high enough, the film which appears in the form of oxide on the terminal surface is crushed. In general it is assumed that in a good connection the contact force should be such that the contact area produced is equal to or greater than the cross sectional area of the wire. In screw connections, crimped connections and wrapped connections the contact area is normally a multiple of the wire cross sectional area. In plug connections, such as on vacuum tube sockets, the contact area is very small. In a Fahnestock clip for example, the contact area is about one quarter of the cross sectional area of the wire.

LIFE

If the electrical resistance of a pressure joint is to remain constant with time, it is the contact area which must remain substantially constant, but not necessarily the contact force. Once the metal particles are tightly interlocked a subsequent reduction in contact force within relatively wide limits does not change the electrical resistance. The resistance will increase only when the force is reduced to such a low value that vibration and handling cause partial separation of the contact area. In such a case two changes may take place:

1. The atmosphere may enter through the fringe of the contact area and a process of corrosion may begin.
2. The effective contact area may be reduced through dislodging some of the contacting particles.

In both cases the resistance is increased. Therefore, to produce a durable connection it is important to have a firm joint and one such that the atmosphere cannot enter the contact area. The term commonly used for such a joint is "gas tight."

ELASTIC RESERVE

The question now arises how much reduction in contact force can be tolerated before a joint loses its gas tightness? In all types of pressure

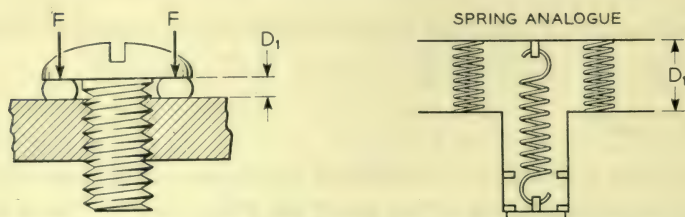


Fig. 2 — Screw tightened and wire compressed (with accompanying "Spring Analogue").

connections, the forces which hold the wire and terminal together are provided by the springiness or elasticity of the materials. The elasticity in a Fahnestock clip is quite apparent because of the long spring member. On the other hand, a screw connection, such as shown in Fig. 2, does not appear to have spring members of any kind. Analysis, however, shows that there is considerable elastic deformation. Most of this elastic deformation is in the elongation of the screw shank and there is also some bending in the screw head and some compression in the screw threads. When the screw is tightened, the wire which is interposed between screw head and nut is also elastically deformed. Since in most electrical connections the wire is a soft material such as copper or aluminum, it is nearly always compressed beyond the yield point and only the recovery of the overstressed material can be considered as elastic reserve.

To determine the usefulness of a connection and compare pressure connections of different kinds, the elastic reserve in the deformed wire and the deformed terminal must be measured or computed. Elastic reserve might be expressed either in terms of stiffness or the potential energy stored in the system. Stiffness S is defined as the ratio of the "applied force F to the elastic return D "; potential or elastic energy E is "one half of the product of the force F and the elastic return D ." ($E = \frac{1}{2}FD$.)

Example: A wire is placed under a screw (Fig. 2) and compressed by the screw head to a thickness D_1 . The screw is then loosened so that it just touches the wire. The wire now has expanded to a certain extent and its new thickness is D_2 . (Fig. 3.) The difference ($D_2 - D_1 = D_w$) is the elastic return and the ratio $F/D_w = S_w$ is the useful stiffness of the wire.

A preferred way of expressing elastic reserve is in terms of stored energy. (Strictly speaking the equation $E = \frac{1}{2}FD$ holds only for springs with constant stiffness. A round wire compressed by a screw head becomes stiffer as the compression increases). The distance, $D_s = D_3 - D_w$, is the elongation of the screw (see Fig. 3). The total energy

stored up is therefore the sum of the energy in the screw and wire ($E = E_s + E_w$).

Screws in terminal blocks are normally made of hard materials such as brass or phosphor bronze. Wires used for the interconnection of components are nearly always of a soft material and have a tendency to creep. If creep takes place in the wire during the many years a screw connection is in use, it is advantageous to have the loss of potential energy in the wire compensated for by the energy stored in the screw.

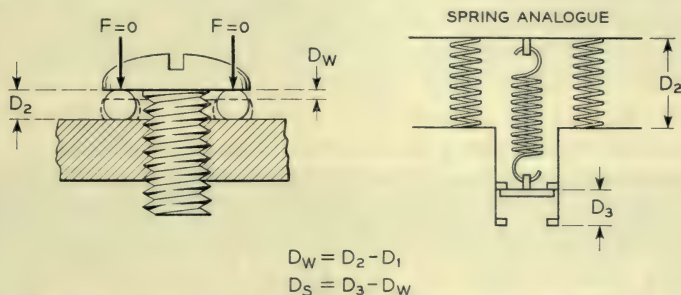


Fig. 3 — Screw loosened and wire not compressed. The recovery of the wire is denoted D_w . The recovery of the screw is $D_3 - D_w$ (with accompanying "Spring Analogue").

A screw, for example, made of soft copper would not be expected to make a lasting connection. If on the other hand the screw is made of a material which has little creep and much elasticity, such as brass or steel, it would act as a spring member and tend to keep the connection tight.

Several typical screw connections were measured to determine the elastic reserve. It was found that on an average the potential energy stored in the screw is about equal to that stored in the wire. Plastic flow of the wire creates an effective bearing area comparable to the area of the screw shank.

THE SOLDERLESS WRAPPED CONNECTION

The detailed analysis of the screw connection as an introduction to the solderless wrapped connection was necessary not only because the screw has such wide use as an electrical pressure connection but chiefly because of its proven value as a durable connection. When new types of pressure connections are put into large scale production, the question invariably arises, What is their life? While considerable analytical work has been done on the cold flow of metals under stress* and while certain

* See Part II.

theoretical predictions can be made on the durability of new connections, it affords additional satisfaction to be able to show that the solderless wrapped connection is in many respects similar in structure and performance to the conventional screw connection. If then this fact is supported by parallel analytical work, there should be little doubt that the solderless wrapped connection is a durable pressure connection.

THE RECTANGULAR TERMINAL

Generally speaking the terminal best suited for a wrapped connection is a terminal of rectangular cross section. It is an inexpensive terminal since it can be blanked from sheet stock or coined from round wire. It is ideally suited for a pressure connection because the edges produce a concentrated high pressure on the wire. The stress distribution in the wire produced by the terminal edges is shown diagrammatically in Figs. 4 and 5. If the wire is wound with high tension around the rectangular terminal, the terminal edges dig into the soft copper wire, crush and shear the oxide on both the wire and the terminal and form a large, intimate and metallurgically clean "gas tight" contact area. An indication of the high pressure is the crushing of the hard nickel silver terminal edge by the soft copper wire. A pattern of contact areas on both wire and terminal is shown in Fig. 6. Several turns of wire are required to preserve the high contact force. In general it is assumed that the first and last two edges around which the wire is wrapped do not contribute much to the joint as contact areas. A seven-turn wrapped connection on a rectangular terminal thus has six effective turns. Each turn contacts four edges or a total of twenty-four contact areas for six effective turns.

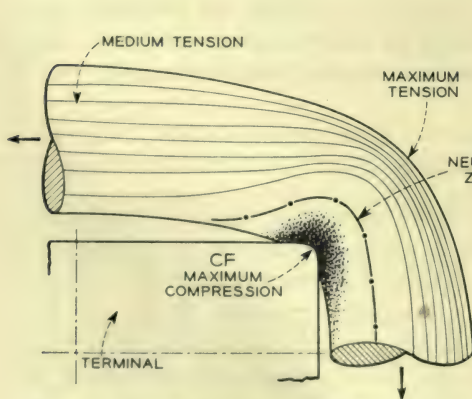


FIG. 4—Stress distribution along one-quarter turn of wire.

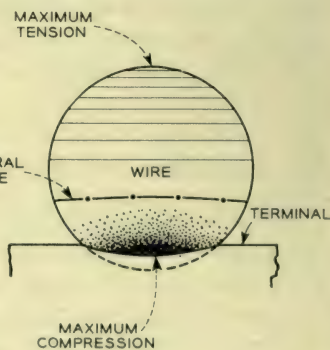
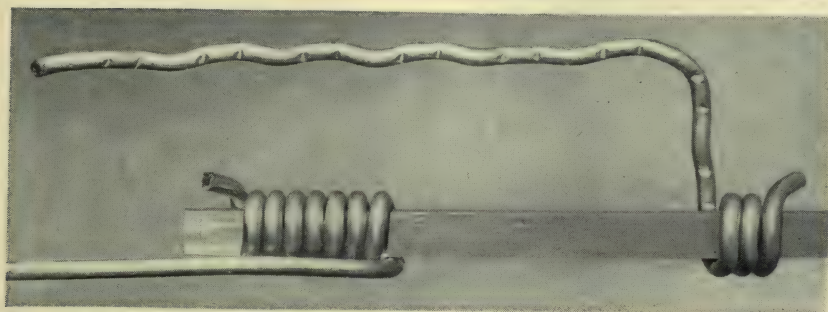
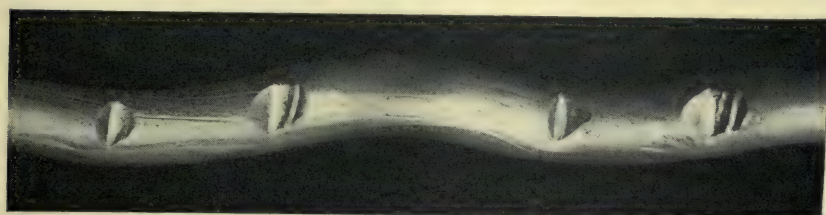


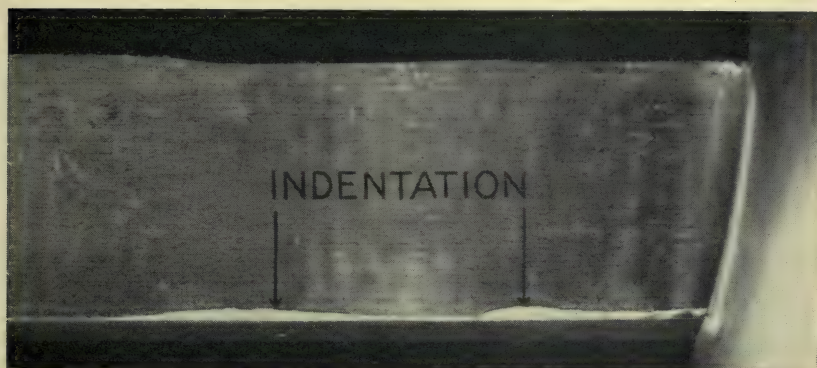
FIG. 5—Cross section through terminal edge showing stress distribution in the wire.



PATTERN OF CONTACT AREAS.



CONTACT AREAS ENLARGED.



THE SOFT COPPER WIRE CRUSHES THE HARD NICKEL SILVER TERMINAL EDGE IN THE WRAPPING PROCESS.

Fig. 6 — Contact areas.

WHAT IS A GOOD CONNECTION?

The quality of a connection depends fundamentally on two factors: the contact area and the contact pressure. As long as there is sufficient pressure and the atmosphere cannot enter the joint, the connection is considered a good one. If, however, the elastic energy which holds the two surfaces together is small, various disturbances may cause a partial separation of the interlocking metal particles and thus effect a change in resistance. For normal telephone applications a good connection may, therefore, be defined as one which not only has sufficient contact area and contact pressure but which also has sufficient elastic reserve to maintain contact area and contact pressure throughout the desired life, which may be forty years or more.

The mechanical disturbances to which a connection may be subjected are: handling, vibration, temperature changes and cold flow.

HANDLING AND VIBRATION

The solderless wrapped connection is well protected from the point of view of handling and vibration. The locking effect on the rectangular terminal or a terminal having well defined edges does not permit loosening of the center turns from the terminal. In vibration tests where conventional soldered connections were compared with solderless wrapped connections, it was found that solderless wrapped connections outlast soldered connections. This is due to the fact that a sudden change in cross section from wire to solder lump localizes the stresses at a very small area. (See Figs. 7 and 8.) In the screw connection a similar condition exists where the wire emerges from under the screw head. In the

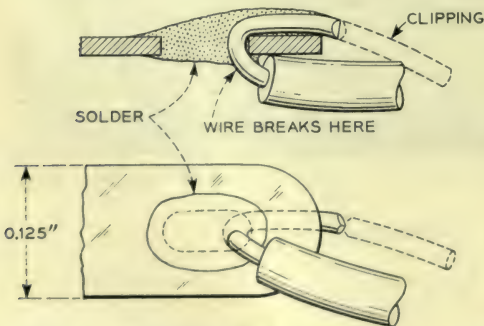


FIG. 7 — Standard solder connection of U-relay.

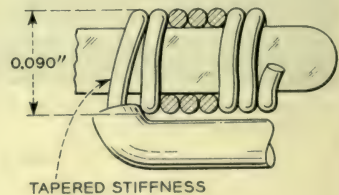


FIG. 8 — Solderless wrapped connection of modified U-relay terminal.

solderless wrapped connection there is no sudden change in cross section and therefore no localization of stresses. The term commonly used to indicate the gradual change in rigidity of the wire as it approaches the anchoring point is "tapered stiffness." (See Fig. 8.)

HEAT AND COLD FLOW

When a pressure connection is subjected to high temperatures, which may be due to large current or to heat transfer from adjacent components, the pressure at the joint is relaxed. This is true in the solderless wrapped connection as well as in the screw connection. The same process of relaxation takes place in normal temperature with time. The relaxation of pressure with temperature and time will be shown in another part of this paper. Under ordinary conditions the relaxation of pressure in a solderless wrapped connection is not sufficiently large to indicate any change in resistance during a forty-year life. Furthermore, as Mason and Osmer point out in their paper, solid state diffusion takes place as time goes on. This process strengthens the joint mechanically and improves it electrically.

QUANTITATIVE EVALUATION OF ELASTIC RESERVE

Because of the above mentioned disturbances to which a pressure connection may be subjected, it is important to know how much elastic reserve is stored in a connection. If no potential energy were stored in the wire and in the terminal, no contact pressure would be produced. If little potential energy were stored in a connection, a slight change in temperature due to differential expansion of the metals would loosen the connection. The same would be true with vibration and handling.

A rough comparison with other pressure connections will serve to illustrate how much elastic reserve a solderless wrapped connection has to have in order to withstand the disturbances to which it may be subjected. The best known pressure connection, and the most universally used, is the screw connection. On a No. 4 screw (0.112"), the force exerted in clamping the No. 24 gauge (0.020") wire is about 135 lbs. The elastic energy is stored by compressing the wire and by elongating the screw shank. Similarly, in a solderless wrapped connection as shown in Fig. 9 a total force of 90 lbs is exerted on the edges of the terminal (24 corners). Here the greater part of the energy is stored in the terminal which receives torsional as well as compressional stress from the tension in the wrapped wire. (See Figs. 10(a) and 10(b)).

Fig. 9 shows in diagrammatic form how the stored energy in a screw connection and a solderless wrapped connection compare. A typical solderless wrapped connection — seven turns of 20-mil copper wire wound with 1300 grams applied force on a 0.0148" x 0.062" nickel silver terminal — has approximately 3 mil pounds of stored energy E . (2.4 mil pounds are stored in the terminal and 0.6 mil pounds in the wire). The screw connection — No. 4-40 screw (0.112") tensioned to 135 lbs on 20-mil copper wire — has approximately 2.7 mil pounds of stored energy E . (Approximately half in the screw and half in the wire). In the screw connection the energy is about equally divided between the screw and the wire whereas in the solderless wrapped connection a larger part of the energy is stored in the terminal. This is advantageous since the hard materials of which terminals are generally made have less cold flow than copper. In a solderless wrapped connection, if the

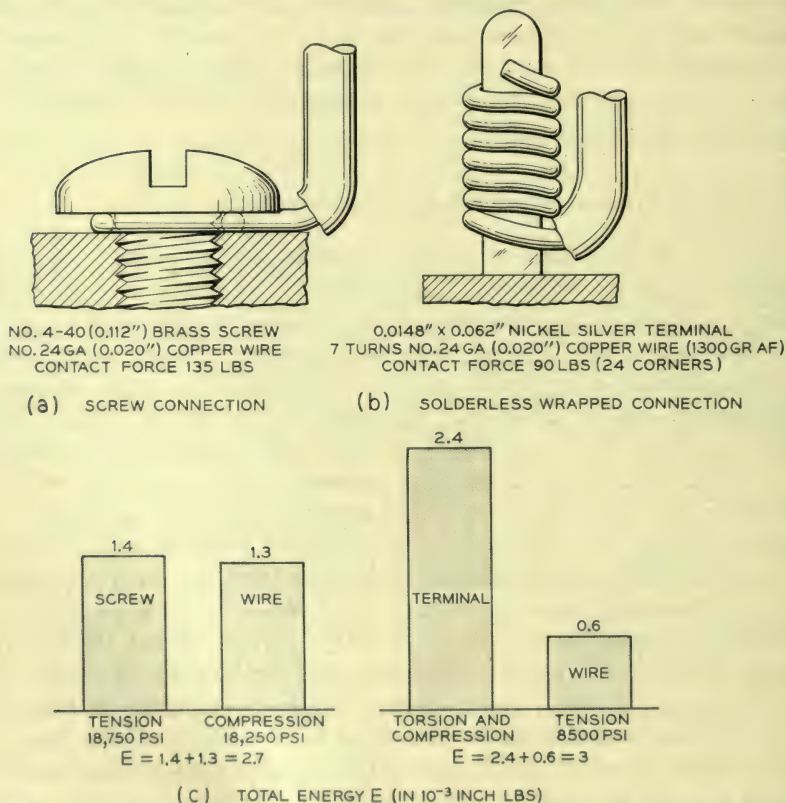


Fig. 9 — Elastic energy stored in screw connection and in solderless wrapped connection.

terminal size is changed to 0.020" x 0.062" there is considerably less energy stored in the terminal and slightly more in the wire. Inasmuch as a screw connection in most cases depends on the human element, that is the amount of torque applied by the operator, it can be expected that some screw connections will be made with a force that may vary from 75 lbs to 150 lbs. The wrapped connection on the other hand, being made with a calibrated tool, can be expected to give substantially the same contact force at all times.

In order to understand more clearly how the wire and the terminal interact when they are under mutual stress and exposed to heat, the elastic deformation of the wire and the terminal must be analyzed. It has been shown in Fig. 4 that the wrapped wire on the four sides of the rectangle is under tension. This tension causes the terminal to twist. If instead of a helix the terminal were surrounded by a series of hoops,

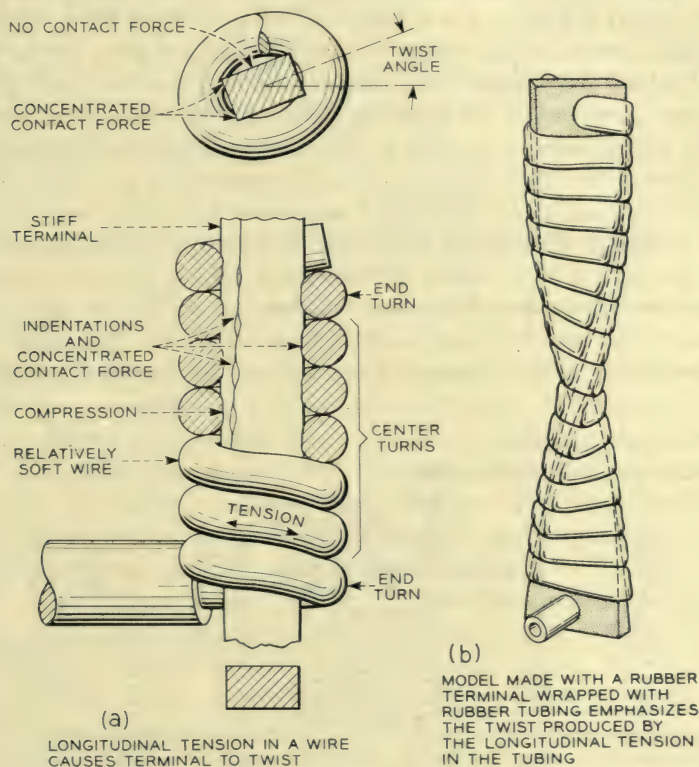


Fig. 10 — (a) Longitudinal tension in wire causes terminal to twist. (b) Model made with a rubber terminal wrapped with rubber tubing emphasizes the twist produced by the longitudinal tension in the tubing.

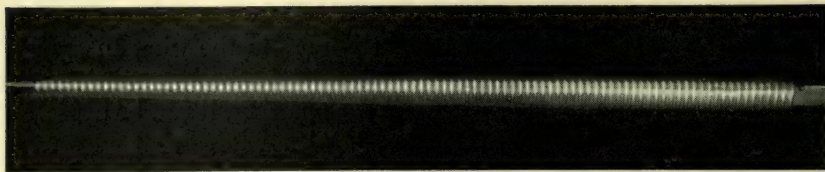


Fig. 11 — Twist of rectangular terminal (100 turns of wire).

the terminal would be compressed at the edges but the terminal would not twist. The terminal twist in the wrapped connection is therefore due to the fact that the terminal is surrounded by a helix and not by hoops. Figs. 10(a) and 10(b) show that a left-hand helix produces a right-hand twist in the terminal. As will be shown later, this visible deformation of the terminal is being used to determine the tension in the wire. The twist in the terminal of a wrapped connection with many turns can readily be seen in Fig. 11. For example an initial twist of 46° is produced in a nickel silver terminal $0.0148'' \times 0.062''$ wrapped with 100 turns of No. 24 ($0.020''$ dia.) copper wire with an applied force of 1300 grams.

One way to visualize the behavior of the wire and the terminal when wrapped under tension, exposed to time and heat and then unwrapped, is to represent the wire and the terminal by linear springs. This is shown schematically in Fig. 12. Position 1 represents both wire and terminal before wrapping. Position 2 represents the wire wrapped on the terminal. Position 3 is the same as Position 2 except that the wrapped terminal has been exposed at room temperature (20°C) for eight days. This causes the terminal twist to relax from 46° to 39° . Positions 2 and 3 are analogous to the wire under tension and the terminal under tor-

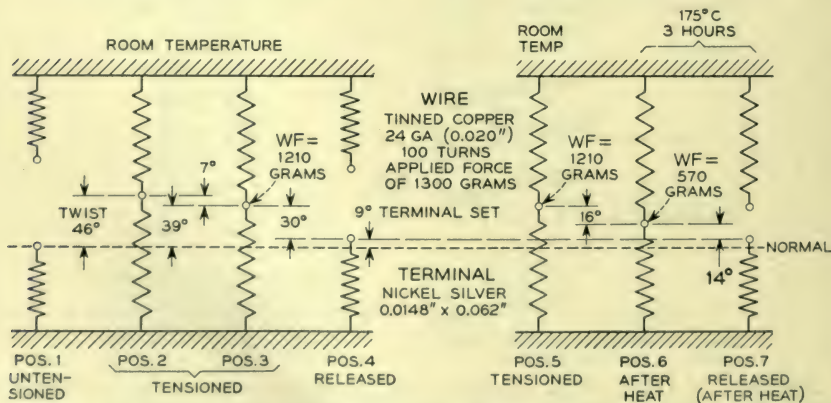


Fig. 12 — Energy in unheated connection proportional to 30° . Energy in heated connection proportional to 14° .

sion. The force WF is the tension in the wrapped wire. This force can be determined by dividing the torque necessary to twist the terminal by the effective moment arm. Since the elongation of the wire cannot readily be measured, the terminal twist was chosen to determine the force exerted at the terminal edge. The 39° terminal twist shown in Posi-

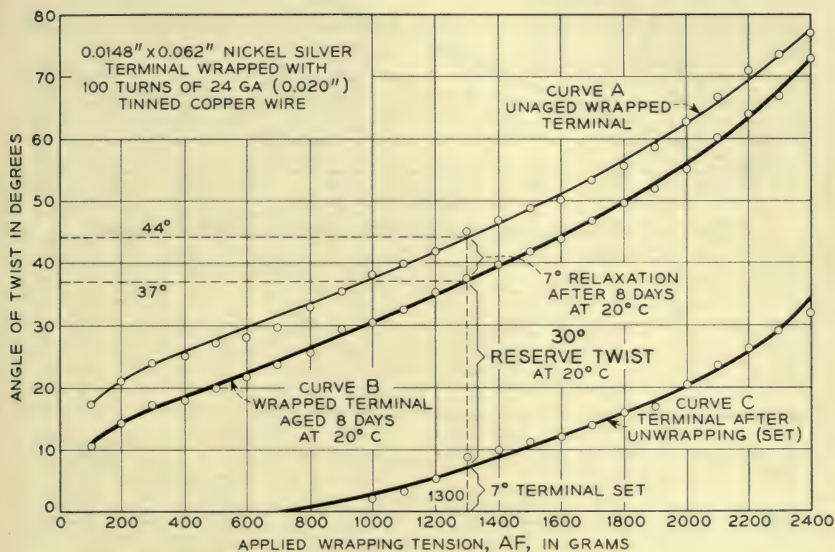


Fig. 13 — Angle of twist in terms of applied wrapping tension.

tion 3, however, cannot be used for determining the force since the terminal may be overstressed as is shown in Position 4. Instead of returning 39° the unwrapped terminal returned only 30°. In other words the terminal has taken a set of 9°. Fig. 12 illustrates the deformation of wire and terminal only for one value of applied tension, namely 1300 grams. If the angle of twist is measured for applied tension ranging from 100 to 2400 grams, a set of curves is obtained as shown in Fig. 13. Curve A shows the angle of twist immediately after the terminal is wrapped. Curve B shows the relaxation after eight days aging at room temperature. Curve C represents the terminal set. The value between Curves B and C is the elastic reserve. For 1300 grams applied tension the elastic reserve is expressed as 30° reserve twist.

Using the before mentioned ratio of torque and moment arm, the force WF can now be determined. The torque required to twist the terminal 30° is 37.2 inch grams. (See Fig. 14.) The effective moment

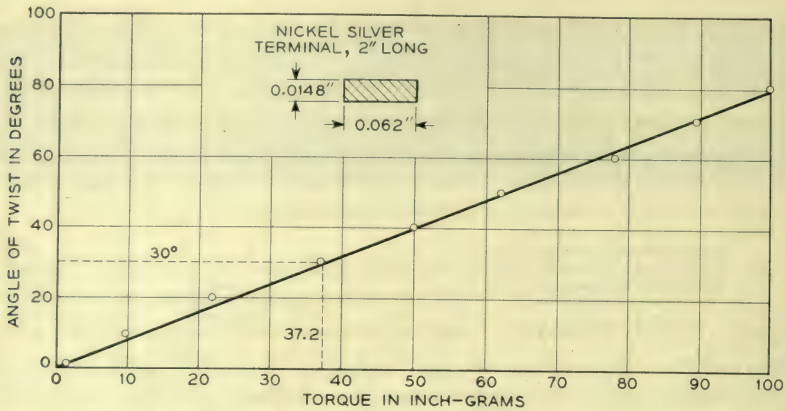


Fig. 14 — Torque required to twist the unwrapped terminal.

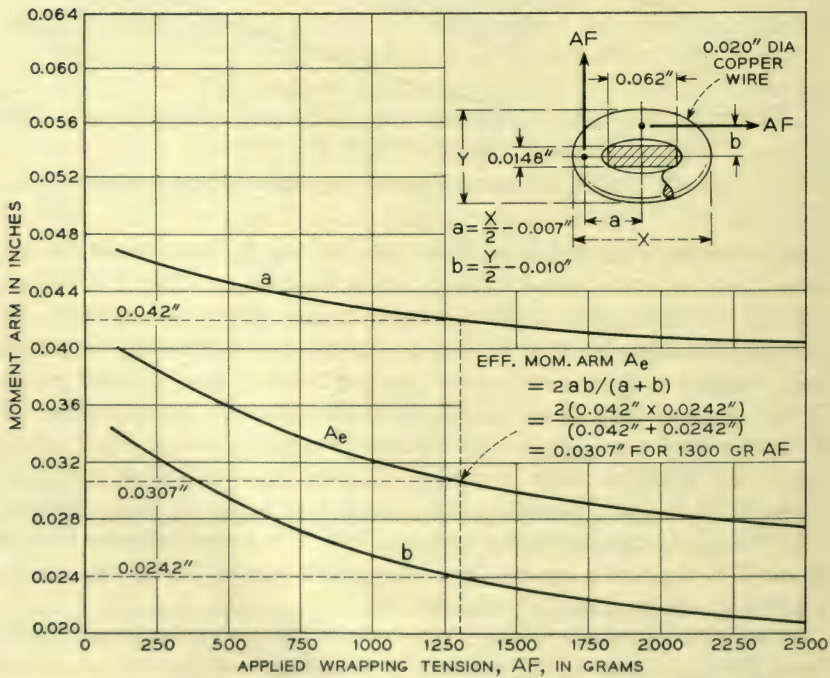


Fig. 15 — Moment arm in terms of tension.

arm A_e , which decreases a slight amount as the wrapping tension AF increases (see Fig. 15), is equal to $2ab/(a + b)^*$. Therefore, $A_e = 2(0.042" \times 0.0242")/(0.042" + 0.0242") = 0.0307$ in. Thus the tension in the wire $WF = T/A_e = 37.2/0.0307 = 1210$ grams. Using the recovery angle of the terminal as a measure of force, the tension in the wire can now be plotted in terms of angular twist. This is shown in Fig. 16. It should be noted that the tension in the wire WF (wrapped

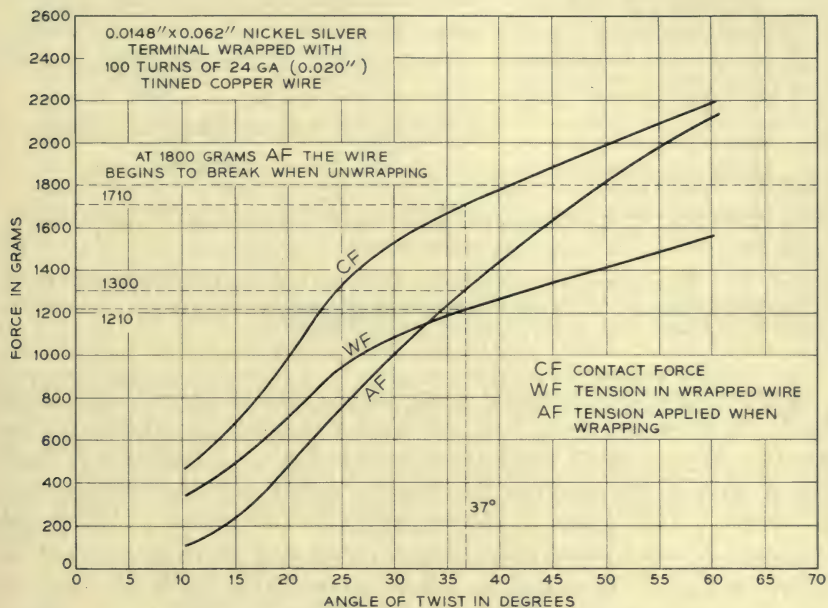


Fig. 16 — Forces in terms of angular twist.

force) is not directly proportional to the wrapping tension AF (applied force). The reason for this is that at low applied wrapping tension the bending of the wire around the corner of the terminal produces an additional increment of tension. For example at an angle of 15° the wrapped tension WF is nearly twice as high as the applied tension AF. The wrapped tension WF and the applied tension AF are about equal when the angle of twist is 33° . At higher values of applied tension, the wrapped tension increases at a much lower rate. This is caused by the terminal taking a set. (See Fig. 13.) At 1300 grams of applied tension, which is the recommended wrapping tension for No. 24 copper wire, the wrapped tension is 1210 grams.

* See Appendix I.

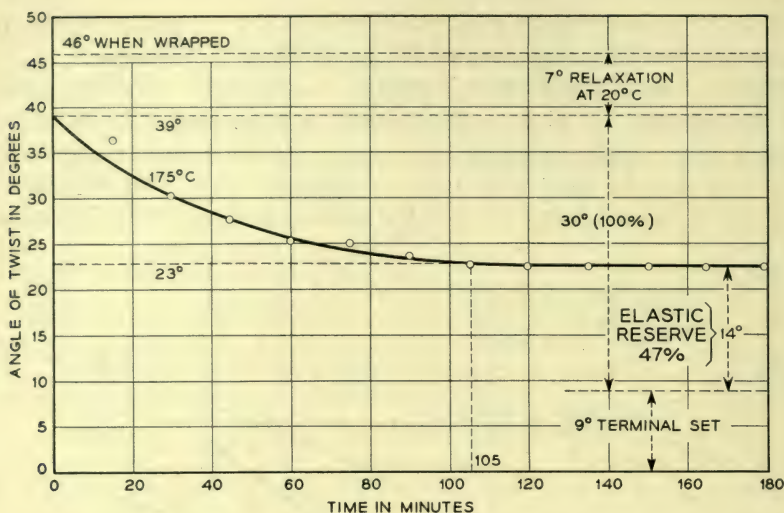


FIG. 17 — Relaxation of wrapped wire after heating for three hours at 175°C. The reserve twist of 14° is an indication of the tension left in the wire which amounts to 47 per cent of the original wrapped tension.

A severe test for a wrapped connection is heating to 175°C for three hours. This relaxes, about half the stress and is considered the equivalent of a 40-year life at 135°F. Position 5 in Fig. 12 is the same as Position 3, that is, the wire has just been wrapped onto the terminal and its tension is 1210 grams. If this connection is now heated to 175°C and the angle of twist noted every fifteen minutes, a curve is obtained as shown in Fig. 17. It should be noted that at 105 minutes the curve is for all practical purposes asymptotic at an angle of 23°. If the heated connection is cooled and unwrapped and the set in the terminal measured, it is found to be 9°. The 14° difference is a measure of the elastic reserve. This is 47 per cent of the wire tension before heating. The corresponding tension WF is then 570 grams. This process is illustrated in positions 6 and 7 of Fig. 12. Position 7 shows that the terminal set of 9° was the same as before heating.

A similar experiment was made with formex insulated wire wrapped on a nickel silver terminal. Instead of subjecting the connection to the heat of an oven, a high current was passed through the wire. Essentially the same curve as shown in Fig. 17 was obtained.

To further check the behavior of springs with complex elastic deformation such as in a wrapped connection on a terminal having edges, measurements were also made with simple helical springs tensioned

within the yield point. Two springs, one of nickel silver wire and the other of copper wire having a stiffness ratio of 5 to 6, were coupled in series (Fig. 18), tensioned to 30 grams and then heated to 173°C for two hours. The tension left after heating was 13 grams or 43 per cent of the original tension. The tension decay curve was similar to that shown for the wrapped connection. (See Fig. 17.) This test shows that in spite of the complex deformation of the wire at the corners of the terminal there is substantial agreement in results of the measurements obtained — namely 43 per cent remaining stress in the case of the helical spring and 47 per cent for the wrapped connection.

STRESSES IN THE FINISHED CONNECTION

Having determined the interacting forces in the solderless wrapped connection, the next questions of primary interest are — what are the stresses in various parts of the connection and what will happen to these stresses in forty years?

Most of the elastic energy stored in the wire is in the portion marked "Medium Tension" (Fig. 4). Here the stress is about 8,500 P.S.I. This assumes that a 20-mil wire is wrapped with 1300 grams applied force. The wrapped force or the useful force obtained from the elastic reserve is then 1210 grams.

The stresses at the corners are not easily determined because they are not uniform. This is shown in Fig. 5. As can be seen, the point of highest

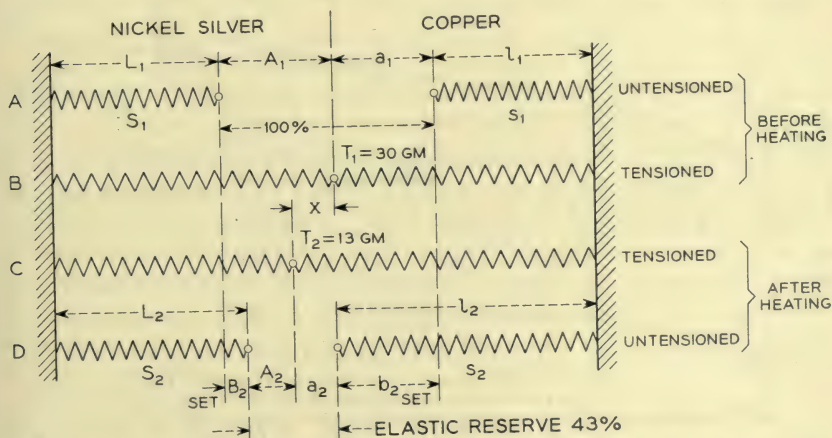


Fig. 18 — Tension T_2 in coupled springs after heating for two hours at 173°C is about 43 per cent of the original tension T_1 . (Tension T_1 approximately 10 per cent below the yield point).

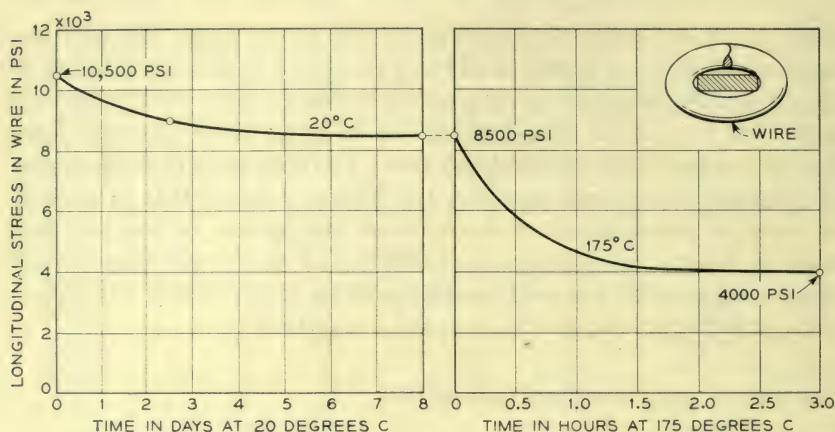


FIG. 19 — Stress relaxation in wire.

concentration is in the center of the contacting area. From this point to the periphery there is a pressure gradient which is similar to that of a circular compressed thin film of viscous material. At the boundary line the pressure is zero. The average pressure within the contact area is about 29,000 P.S.I. but the maximum stress in the center of the contact area may be as high as 100,000 P.S.I. The relaxation, that takes place in eight days at room temperature (Fig. 13), is assumed to be due to the very high initial stress in the center of the contact area which seeks equalization.

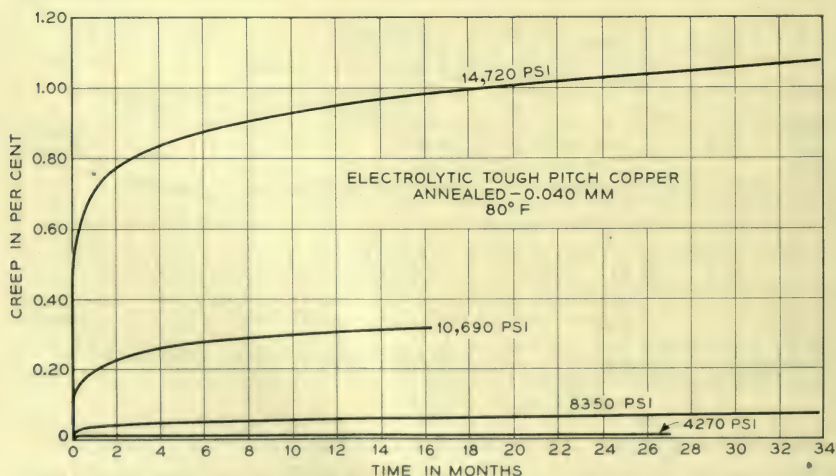


Fig. 20 — Creep curves of annealed copper for various stresses. (Courtesy of Chase Brass and Copper Co., Waterbury, Conn.).

Summarizing the stresses in the connection, one may therefore say that in the portion of the wire where most of the elastic energy resides, the stress after eight days is about 8,500 P.S.I. and at the points of contact 29,000 P.S.I. After forty years these stresses will be approximately 4,000 and 13,500 P.S.I., respectively or 47 per cent of the original stress. (See Fig. 19). As may be seen in the creep curves shown in Fig. 20, a stress of 8350 P.S.I. reaches a creep value of about 0.07 per cent in three years and from then on for all practical purposes ceases to creep.

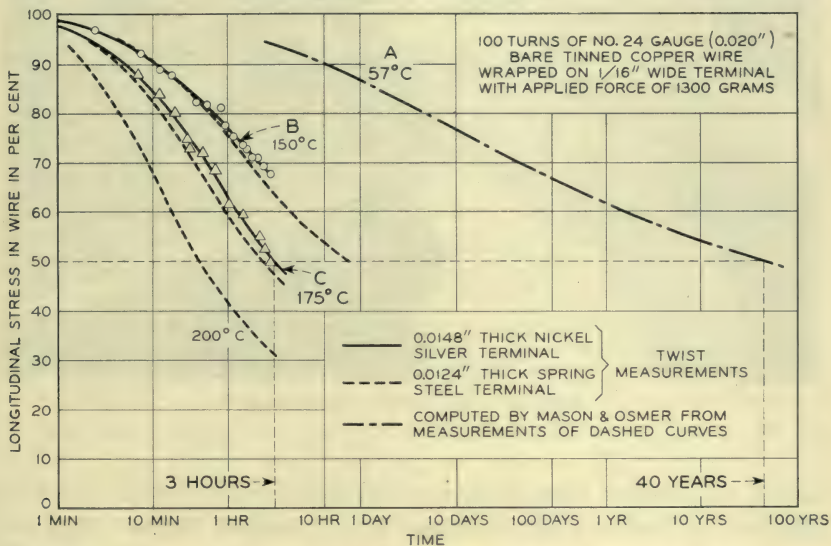


FIG. 21 — Stress relaxation in wire plotted on a logarithmic time scale.

The effect of time and temperature on the longitudinal stress in the wire can better be seen by curves plotted on a logarithmic time scale (Fig. 21). The initial stress after transient relief of three days is considered as 100 per cent. Curve A shows that at a temperature of 57°C (135°F) — which is the maximum temperature that solderless wrapped connections will be subjected to — the longitudinal stress relaxes approximately 50 per cent in 40 years. To reach the 50 per cent value at a temperature of 175°C takes approximately three hours (Curve C). Curves B and C show that the relaxation in the wire is essentially the same for either nickel silver or spring steel terminals.

METHOD OF WRAPPING

In nearly all soldered connections where a wire is to be joined to a terminal the procedure is as follows: The operator takes the skinned

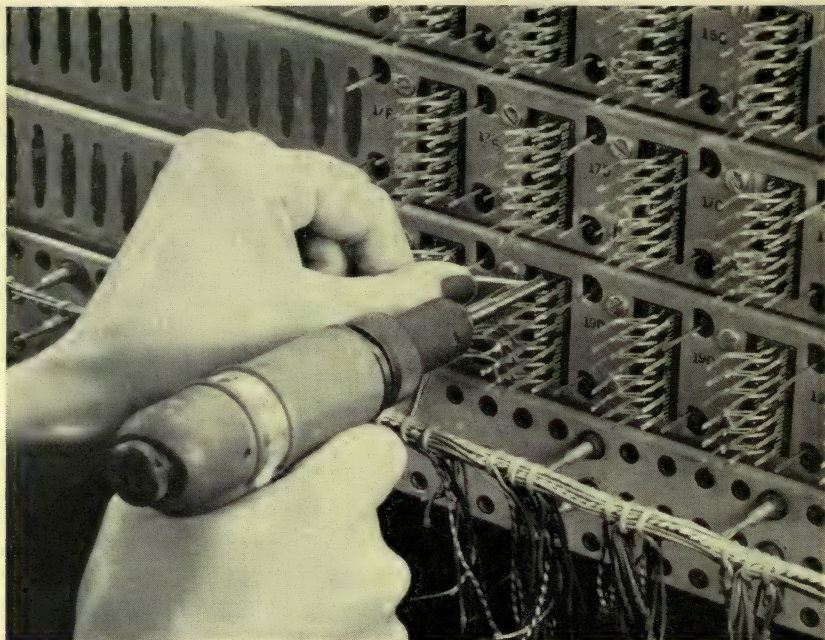


Fig. 22 — Air-driven wiring tool.

end of an insulated wire, hooks or threads it onto the terminal and applies solder. The hooking or threading is important because it is customary in production to attach several wires at first and then solder. The question now is: If a similar procedure is to be followed with a wrapped connection where the wire must surround the terminal with a high pressure, how do we produce that pressure?

In a screw connection a force is obtained by a high lever ratio. In a crimped connection the force is applied by heavy and powerful compression tools. These tools are not suitable for connecting wires to closely spaced terminals such as shown in Fig. 22. To produce high tension in the wire while it is being wrapped onto the terminal, a new method of tensioning had to be devised.

In the manufacture of helical springs it is customary to anchor the end of the wire in a hole in the arbor and tension the wire with a friction pad. By rotating the arbor a helical spring is produced. For closely spaced terminals this method is not practical as the wire cannot be fed tangentially to the terminal and the terminal cannot be rotated. A new

wire connecting concept was proposed whereby a rotating spindle housed a stationary terminal in an axial opening in the spindle and was provided with a second opening radially separated from the axial opening and arranged to accommodate a wire. When the spindle was rotated the wire was caused to form a spiral about the stationary terminal. One method involved anchoring the wire in the second opening and feeding the wire tangentially to the terminal as the spindle was rotated. Due to certain limitations inherent in tangential feed onto a stationary terminal an improved method was finally chosen. This is the axial feed method which is particularly adapted to wrapping closely spaced terminals of all cross sections. The operation of loading the wire and wrapping the connection is shown in Fig. 23. Position A shows the tool tip, Position B the bare wire 2 inserted into the feed slot 4, Position C the anchoring of the wire by bending it into the notch 5, Position D the terminal insertion and Position E the wrapping of the wire 2 by rotating the spindle 1 around the terminal 3. Position F is the finished connection. A more detailed drawing of the tool tip is shown in Fig. 24.

WRAPPING TENSION

The tension in the wire is produced by rotating the spindle 1 around the terminal 3 (Fig. 24) thus pulling the short skinner wire 2 out of the feed slot 4. In the process of pulling the wire out of the slot and wrapping it around the terminal each increment of the skinner wire length undergoes several bending operations. The first bending occurs at the edge R of feed slot 4 where the wire is bent through an angle of less than 90° . The second bending is the straightening out operation of the bent wire. The third bending takes place as the wire is wrapped around the terminal. All three bending processes contribute to the tension with which the wire is wrapped. The dimensions which control the tension and are therefore of engineering importance are the radius R at the tool tip (See Figs. 25 and 25(a)) and the wall thickness W (Fig. 24). The bending forces are inversely proportional to the respective bending curvatures and the frictional forces in turn are proportional to the bending forces. The tension imparted into the wire as it is wrapped around the terminal, however, is not only due to the friction alone but to the combined effect of friction and bending effort. If the wire were completely elastic and the friction zero, no tension could be produced. But there would be tension in the wire if the friction were zero and the wire only partly elastic such as copper wire. There also would be tension if a completely elastic wire would be pulled around an edge having friction.

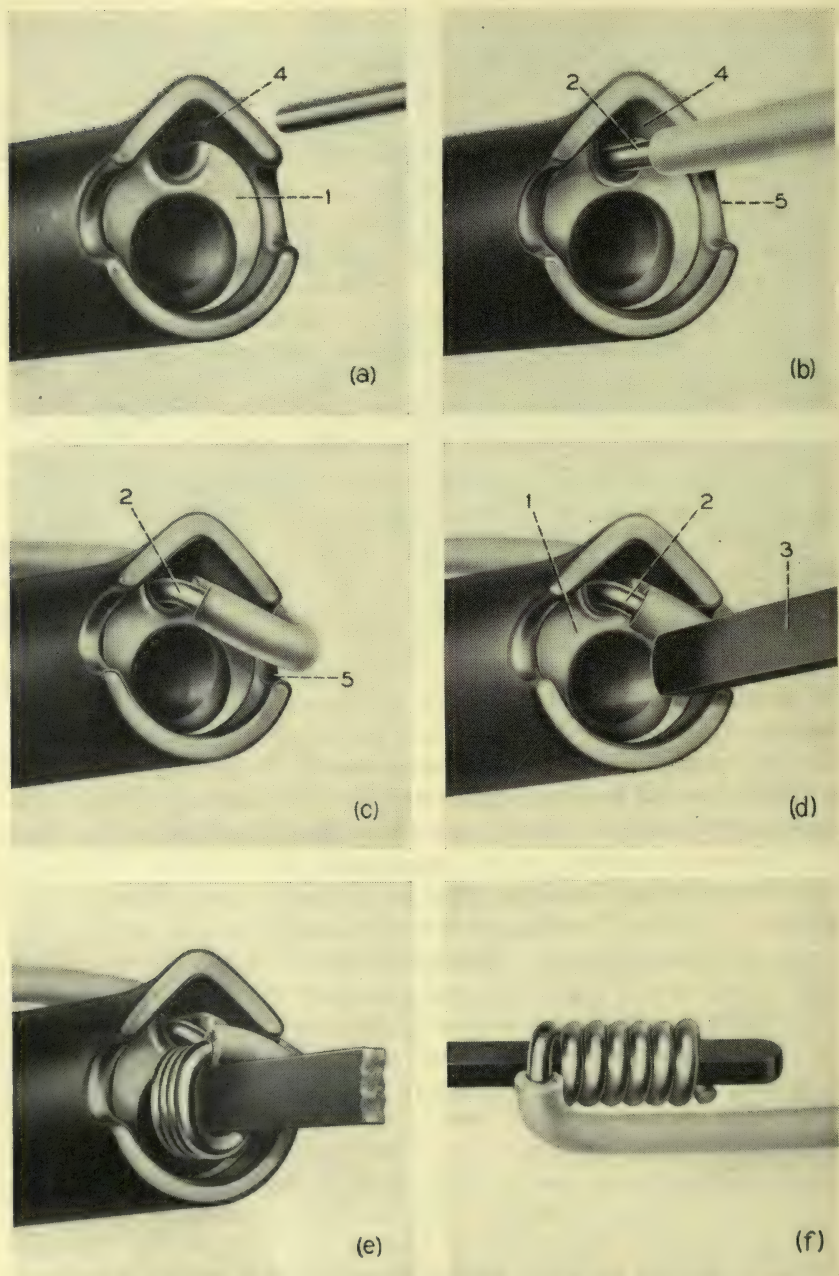


FIG. 23 — The wrapping process. A — Tool Tip. B — Wire Inserted. C — Wire Anchored. D — Terminal Insertion. E — Wire Wrapped. F — Finished Connection.

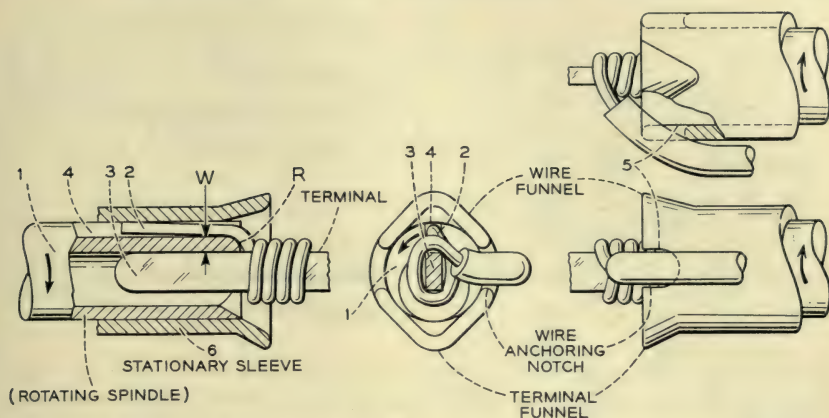


Fig. 24 — Method of wrapping the skinner wire on a rectangular terminal.

CLOSELY SPACED TERMINALS

When the terminals are closely spaced the stationary sleeve 6 and the anchoring notch 5 are used in order to anchor the first turn of wire to the terminal. (See Fig. 26.) However, when the terminals are not closely spaced the sleeve and notch are desirable but are not necessary since the insulated portion of the wire can be held by some means external of the tool at an angle of approximately 90° with respect to the tool spindle. The high acceleration of the wrapping motor produces a mass reaction of the wire leading up to the terminal. This counterforce coupled with a slight tension of the supply wire applied by the external

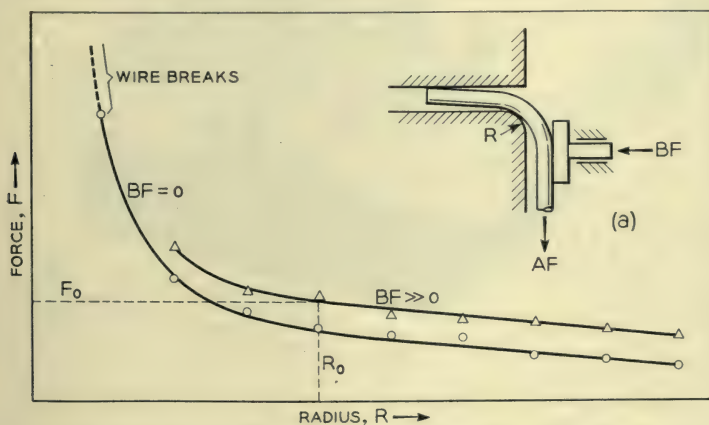


Fig. 25 — Wrapping tension is controlled by edge radius.

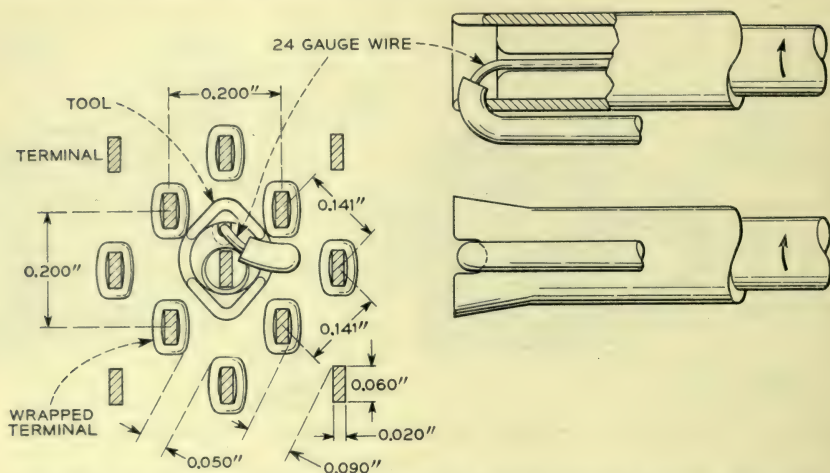


Fig. 26 — Space occupied by wrapping tool between closely-spaced terminals.

wire guiding means is sufficient to insure wrapping of the first turn. The following turns need no further anchor as the first turn locks the wire to the terminal.

The trend toward making circuit components smaller is now marked in all branches of communication engineering. With a tool tip such as

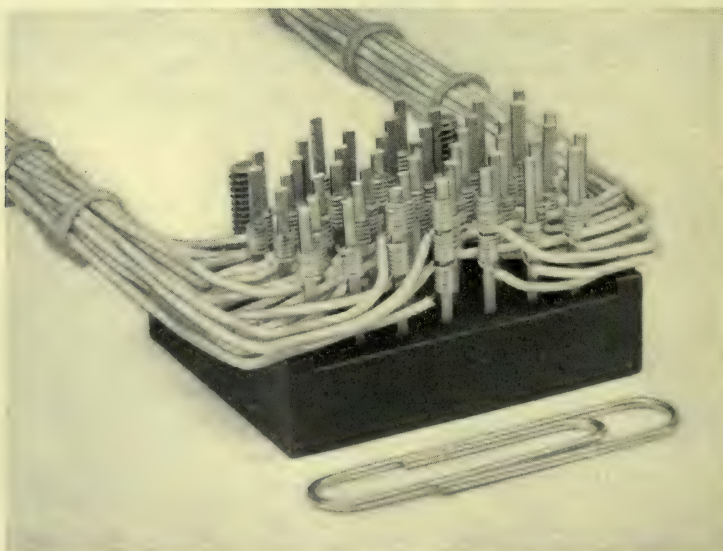


Fig. 27 — Forty-four point terminal block, 132 connection capacity, only 48 connections of 26 gauge (0.0159") wire shown. Occupies $1\frac{1}{8}" \times \frac{3}{4}" \times \frac{9}{16}"$ or $\frac{1}{2}$ cu in of space.

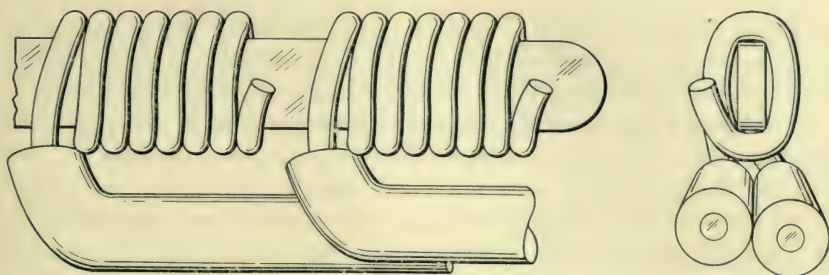


Fig. 28 — Double connection.

shown in Fig. 24, it is possible to wire apparatus having a terminal spacing as close as $2\frac{1}{2}$ times the terminal width. (See Fig. 26.) A terminal block $1\frac{1}{8}$ " by $\frac{3}{4}$ " by $\frac{9}{16}$ " having 44 terminals is shown in Fig. 27. The cables shown contain forty-eight No. 26 (0.0159" dia.) wires all wrapped on the terminals. Each terminal is capable of accommodating three wires or a total of 132 connections may be made in an area of less than one square inch. An enlarged view of a double connection is shown in Fig. 28.

REMOVAL OF CONNECTION

The solderless wrapped connection may be removed from its terminal by two methods. The most convenient method is by stripping. Two types of tools may be used for this purpose. The specially formed jaws of a pair of pliers are hooked in the back of the connection as shown in Fig. 29. By applying a force the connection may be stripped off. The other tool for stripping is shown in Fig. 30. The stripping force varies with the tightness of the wrapping and is plotted in terms of applied

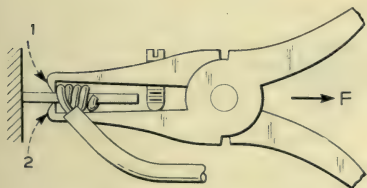


FIG. 29 — Stripping of solderless wrapped connection.

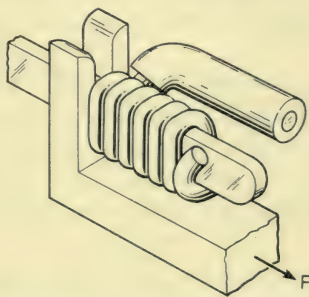


FIG. 30 — Stripping a solderless wrapped connection from a terminal.

wrapping tension in Fig. 31. Another method of removing a connection is by unwinding the helix. This may be done by using a pair of pliers as shown in Fig. 32. Either end of the wire may be used for unwrapping.

A terminal is not seriously damaged by stripping off a wrapped wire, however, the re-use of the stripped off wire is not recommended. A wire may be reconnected by skinning to the proper length and wrapping. When the wire is not sufficiently long to provide the necessary number of turns to insure a good connection, one or two turns may be wrapped and then soldered.

CONNECTION OF LARGE AND SMALL WIRES

There is no upper limit to the size of wire wrapped on adequately proportioned terminals. Connections have been made with both aluminum and copper wire over 200 mils in diameter with satisfactory results. The torque necessary to wrap large wire is considerable, since it increases with the third power of the diameter. A 20-mil wire requires a winding torque of 100 inch grams whereas, a 200-mil wire requires 100,000 inch grams (18 foot pounds). Wires as small as No. 39 (0.0035" dia.) may also be wrapped, however, the design of the wrapping tool must be changed slightly in order to facilitate the loading of the fine wire into the tool.

DIMENSIONAL RELATIONS

The data given in this paper refer only to No. 24 copper wire 20 mils in diameter. The terminal width most frequently used in conjunction with this size wire is about one-sixteenth inch or three times the wire diameter. The terminal width may also be twice the wire diameter or slightly less, however, the one-sixteenth inch size has been chosen for

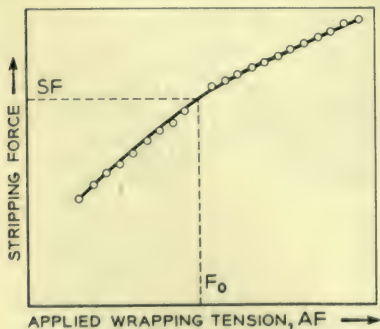


FIG. 31 — Stripping force in terms of applied wrapping tension.

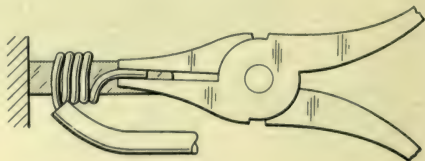


FIG. 32 — Unwinding wrapped connection with pliers.

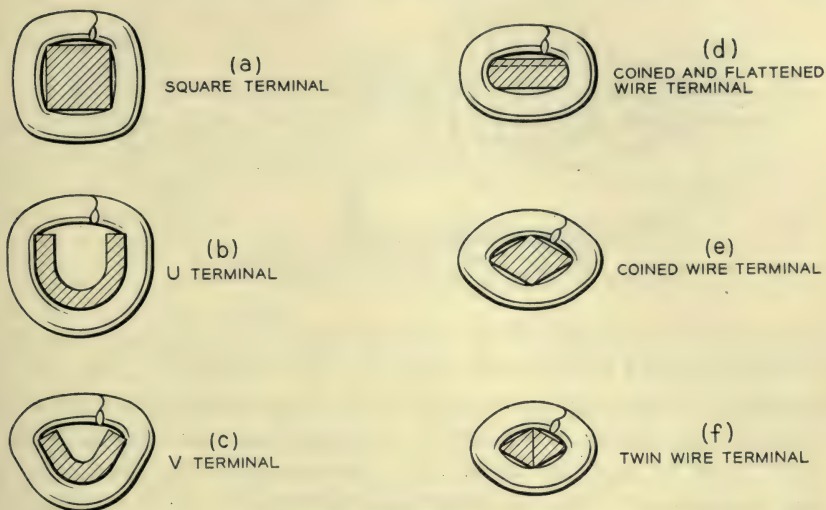


Fig. 33 — Various terminals.

good visibility. When smaller wires are used, the tendency is to make the terminal width greater than three times the wire diameter for better visibility. The terminal thickness depends to a great extent on the shape of the terminal. For a rectangular terminal the thickness may vary from three times to one-half the wire diameter. When the terminal thickness is less than one-half the wire diameter, the terminal may twist too much during the wrapping operation.

TYPES OF WRAPPED CONNECTIONS

The rectangular terminal is not the only terminal which lends itself to a good solderless wrapped connection. Any terminal offering one or more contacting edges substantially crosswise to the axis of the wrapped wire will make a good connection.

Since rectangular terminals of very thin material may twist excessively during the wrapping process the preferred shape is a U or V as shown in Fig. 33. These terminals are capable of storing even more elastic energy than a rectangular terminal of equal cross sectional area. The U and V terminals are particularly suited for vacuum tube sockets and thin relay springs.

Flattened or coined single wires as well as coined twin wires may be used as terminals for solderless wrapped connections. These are shown in Fig. 33.

Stranded wire connections have been made by laying the strands

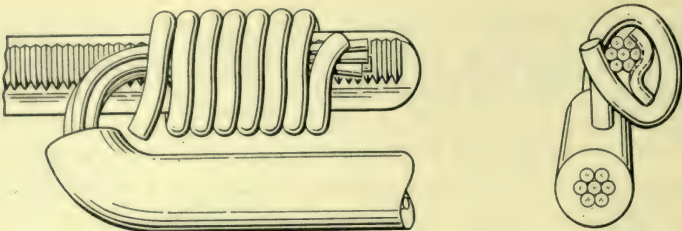


Fig. 34 — Stranded wire connection.

along the serrated edge of the terminal and then wrapping both the terminal and the strands with solid wire. This is shown in Fig. 34. A stranded wire may also be wrapped in the same manner as a solid wire. However, the strands of the wire to be wrapped must be dipped in pure tin in order to bind the strands into the equivalent of a solid wire. A test was made on a No. 24 (0.020" dia.) wire having seven strands and three twists per inch. The preliminary resistance and aging tests have shown connections of this type to be good.

Enameled wire also has been wrapped. The tool for this purpose has a combination edge at the wrapping tip. Part of the edge has an arc for producing the wrapping tension and the other part has a scraping edge for removing the enamel from the underside of the wire as it is being wrapped onto the terminal.

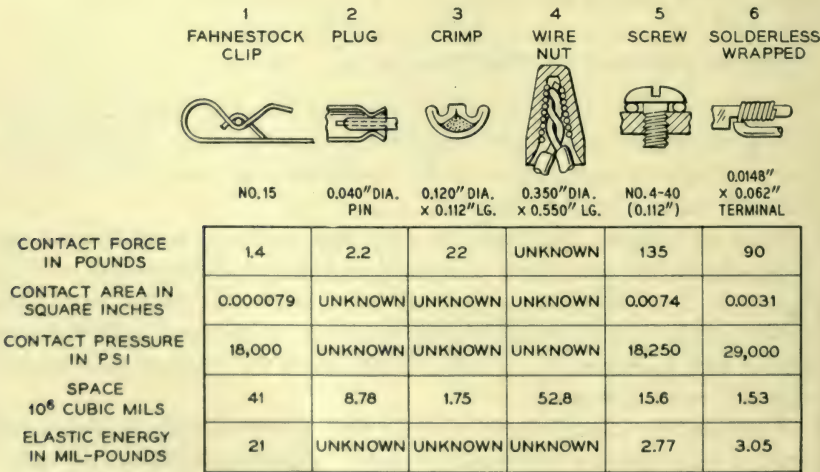


Fig. 35 — Comparison of pressure connections for No. 24 (0.020") wire.

EVALUATION

The solderless wrapped connection has been compared with other pressure connections. (See Fig. 35.) However, when compared with a soldered connection its advantages are as follows:

1. A substantial reduction in wiring defects in manufacture and in service because of:—

- a. Greater uniformity obtained with a calibrated tool.
- b. Less breakage of wires due to handling and vibration.
- c. No solder splashes.
- d. No clippings.
- e. No cold joints.
- f. No rosin joints.

2. Less expensive connection.
3. More compact connection.
4. More clearance between current carrying parts.
5. Easy to disconnect.
6. Saving of tin — a critical material.
7. No contact contamination from soldering fumes.
8. No damage to heat sensitive materials in circuit components.
9. No hazard from hot soldering iron.

SUMMARY

A good pressure connection depends on the amount of elastic energy which can be stored in the mutually stressed members, namely the wire and terminal. If the ratio of elastic energy to the size of the connecting members is favorable, and the contacting areas are sufficiently large, then the connection can be termed good. The solderless wrapped connection when properly proportioned not only meets these requirements, but is uniform in quality and low in cost.

APPENDIX I

EFFECTIVE MOMENT ARM FOR TORSION OF RECTANGULAR TERMINAL

In this appendix the relationship between the wrapped tension in the wire and the twist of the terminal will be analyzed. The structure of the solderless wrapped connection is equivalent to a terminal having springs attached between its edges as shown in Fig. 36. The springs are arranged in such a way as to form a helix of pitch p . Now let:

S_u = torsional stiffness of a unit length terminal

WF = wrapped tension

$2a$ and $2b$ be the sides of the rectangular cross section of the terminal.

ℓ_1 = lead of wire helix on long side ($2a$).

ℓ_2 = lead of wire helix on short side ($2b$).

N = number of turns of wire

The classical formula for Torsion is

$$\theta = \frac{\text{Torque} \times \text{Length}}{S_u} \quad (1)$$

For an equivalent "spring" attached to the long side the axial length of the terminal is ℓ_1 , the torque is WFb^* and hence the deflection $WFb\ell_1/S_u$; for the short side the length is ℓ_2 , the torque is WFa^* and the deflection $WFa\ell_2/S_u$. For a complete turn of the helix the length of the terminal is therefore, $2(\ell_1 + \ell_2) = p$, and the torsional deflection is

$$\delta = \frac{2WF}{S_u} (b\ell_1 + a\ell_2) \quad (2)$$

It is logical to assume that

$$\frac{\ell_1}{\ell_2} = \frac{a}{b} \quad (3)$$

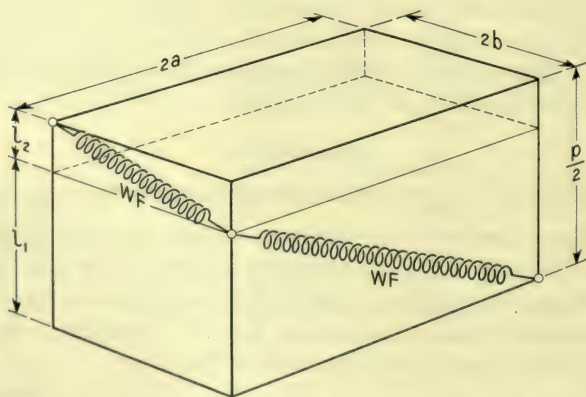


Fig. 36 — Equivalent structure of wrapped connection (only half turn shown).

* This is approximate and holds closely for $\ell_1 \ll 2a$ and $\ell_2 \ll 2b$

Substituting (3) into (2), the deflection per turn becomes

$$\delta = \frac{2ab}{a+b} p \frac{WF}{S_u} \quad (4)$$

Since all effective turns are similar the total deflection Δ for N turns is

$$\Delta = WF \left[\frac{2ab}{a+b} \right] \frac{(pN)}{S_u} \quad (5)$$

pN is the total length of the terminal. By equation (1) the effective torque must then be

$$\text{Torque} = WF \left[\frac{2ab}{a+b} \right] \quad (6)$$

and hence the effective moment arm, A_e

$$A_e = \left[\frac{2ab}{a+b} \right] \quad (7)$$

Solderless Wrapped Connections

PART II — NECESSARY CONDITIONS FOR OBTAINING A PERMANENT CONNECTION

By W. P. MASON and T. F. OSMER

(Manuscript received February 9, 1953)

In order to study the stresses and strains occurring in a solderless wrapped connection, a photoelastic technique using photoelastic bakelite and a photo-plastic technique using polyethylene have been used. Polyethylene has a stress strain curve similar to a metal and can be used to investigate strains in the plastic region. Using these techniques, it is shown that the connection is held together by the hoop stress in the wrapping wire. In order to lock this in, a dissymmetry from a circular form has to occur. This may be in the direction of an oval shape or a square or rectangular shape. Sharp corners are preferred since a more definite contact area results. A number of rules are derived for constructing the most satisfactory solderless wrapped connection.

It is shown that the connection between the wire and terminal is intimate enough to permit solid state diffusion, but the strains are not high enough to cause cold welding of the connection. The life of the joint depends on the twin processes of stress relaxation and self diffusion. Stress relaxation occurs at a rate such that half the hoop stress is relaxed in 2500 years at room temperature. This loss of stress is compensated by the diffusion of one part of the joint into the other. Since the activation energies for stress relaxation and self diffusion are approximately equal for most metals, the two effects complement each other and produce a connection which should remain unchanged for times in excess of forty years under any likely ambient conditions.

INTRODUCTION

The solderless wrapped connection described in the paper by R. F. Mallina (see page 525) provides a very satisfactory and economical method for making connections with apparatus terminals when such

connections are properly made. Not all methods of wrapping or all types of terminals are equally satisfactory and it is the purpose of this paper to describe investigations that have been made to determine the necessary conditions for the best wrapped terminal. These investigations include a photoelastic investigation of the stresses in the terminal and a photoplastic investigation of the strains in the outside wrapping wire. A new photoplastic material, polyethylene, has been used which has a stress strain curve similar to a metal and a birefringence proportional to the strain. The use of this material makes possible the evaluation of strains in the plastic region and may find applications in other plastic flow problems such as the extrusion of metals.

Even after such terminals have been satisfactorily made, there remains the question of whether they will have sufficient life to satisfy the requirements of the telephone plant. A design objective for most relays and other switching apparatus of the telephone plant is an uninterrupted trouble-free life of forty years. Hence, unless the connections are to be the limiting factor in the maintenance of the equipment, they also should have a minimum life of forty years under the conditions for which the apparatus is designed. In order to investigate the probable length of life of such connections, theoretical and experimental work on stress relaxation in metals has served as the basis for calculations and tests. These have been extended to the materials and conditions of the wrapped solderless connection and the results indicate that the life should be adequate even under very severe ambient conditions.

PHOTOELASTIC ANALYSIS OF STRAINS IN TERMINALS OF THE SOLDERLESS WRAPPED CONNECTION

In studying the conditions necessary to insure a good solderless wrapped connection, it is desirable to know what strains occur in the terminals and in the wrapping wires and how these vary with the terminal shape, the winding force and other variables entering into the construction of the connection. While some of these strains can be surmised from the winding conditions and the shape of the terminals, it is difficult to obtain any quantitative results by calculations on account of the fact that the desirable terminal shapes are rather complicated and because a good many of the strains are in the plastic region.

To remedy this difficulty, use has been made of a photoelastic and a photoplastic technique. For the inside terminal, all the strains, except at the corners where the wires make contact with the terminals, are elastic and can be approximated with an ordinary photoelastic tech-

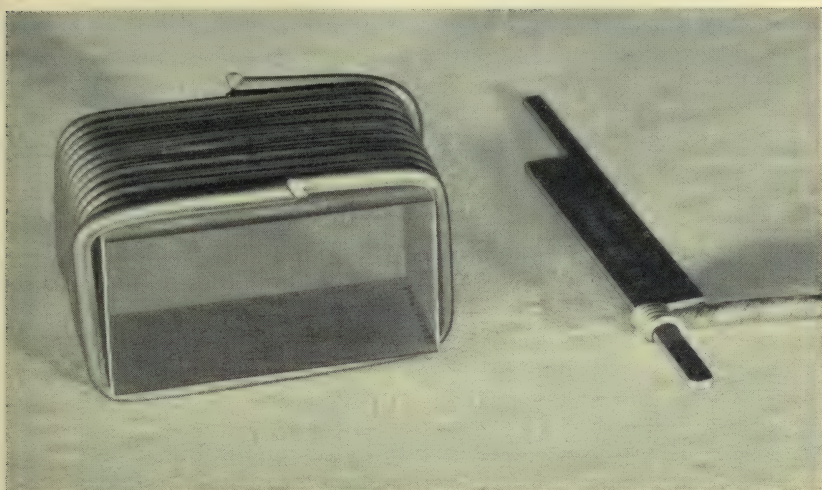


Fig. 1 — Photograph of photoelastic model and solderless wrapped connection.

nique using photoelastic bakelite. Fig. 1 shows one of the photoelastic models as compared with the metal solderless wrapped connection that it simulates. While the ratios of the wire diameter to the terminal dimensions are different in the model from those in the connection, considerable information can be obtained about strains in the terminals and wires from the photoelastic model.

The wrapping gun described in the previous paper puts a tension on the wire. To simulate the tension, the photoelastic model is placed in a chuck, one end of the wire is anchored to the chuck and an appropriate weight is applied to the other end of the wire. The specimen is then rotated by the chuck and a definite number of turns of copper wire are wound around the specimen. The extra wire is then clipped off and it is found that the wire tightly adheres to the terminal in the manner of a metal solderless wrapped connection. The specimen is then polished on the two ends up to the end wires and is put into the polariscope of a photoelastic analyzer. For obtaining the isochromatic lines, i.e., the lines occurring when the ordinary and extraordinary rays differ in path length by a half wavelength or some multiple of a half wavelength, the elements of the polariscope, as shown by Fig. 2, contain a quarter-wave plate before and after the specimen. These have the effect of making the plane wave from the polarizer circularly polarized and eliminate the isoclinic lines which mark the directions of the slow and fast axes of the material. Fig. 3 shows the isochromatic lines for a square

specimen 0.4 inches square wound with nineteen turns of 0.050-inch copper wire under a constant weight of twenty-eight pounds, which is a stress of 14,300 pounds per square inch. It is evident from the sharpness and number of the lines that can be seen that we are dealing with a case of plane stress that can be analyzed by the method discussed in the Appendix. Since the stress strain curve of copper wire has the form shown by Fig. 4 with a yield stress* of about 26,000 pounds per square inch and a breaking stress of about 34,000 pounds per square inch, the applied winding stress is about 42 per cent of the breaking stress. Fig. 4 shows also the recovery measured for copper wire. The recovery curves are quite accurately parallel to each other, but the larger the strain the smaller the percentage recovery. Using the isoclinic lines shown by Fig. 24 of the Appendix and the method of analysis discussed there, the stresses across and perpendicular to the line of "eyes" of Fig. 3 are shown by Fig. 5. The stress perpendicular to the line of eyes measures the total compressive stress put on the terminal by the hoop stress in the wire and from this measurement the average hoop stress remaining in the wire can be calculated as follows. The cross section of which

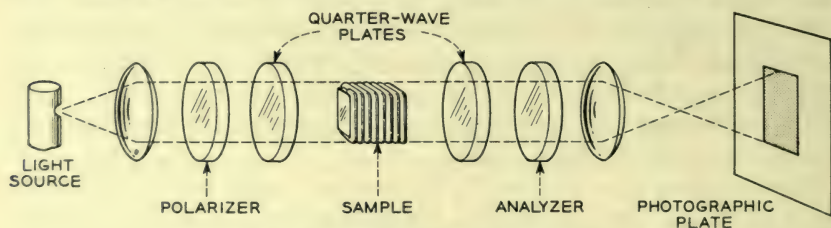


Fig. 2 — Elements of polariscope.

this force is applied is the width 0.4 inches by the length of the specimen 0.95 inches and hence the force applied by all the turns is

$$F = 0.95 \times 0.40 \times 2000 = 760 \text{ pounds.} \quad (1)$$

Since there were nineteen turns of wire wound around the specimen and each turn has two sides exerting a tension on the bakelite, the average tension remaining in the wire, required to balance the compressive stress, is

$$T = \frac{760}{2 \times 19} = 20 \text{ pounds,} \quad (2)$$

* In this paper the yield stress is taken as the point of greatest curvature of the stress-strain curve.

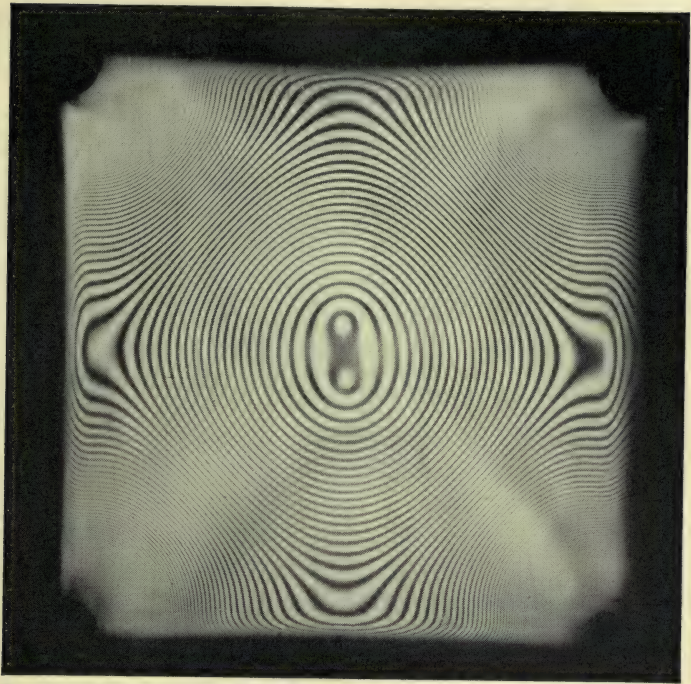


Fig. 3 — Isochromatic lines for a square model 0.4 inches on a side wrapped with nineteen turns of 0.050 inch copper wire with a constant load of 28 pounds.

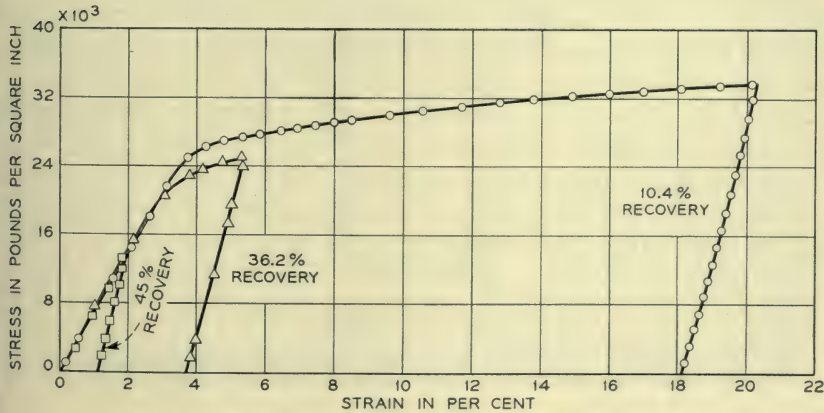


Fig. 4 — Stress-strain recovery curves for copper wire.

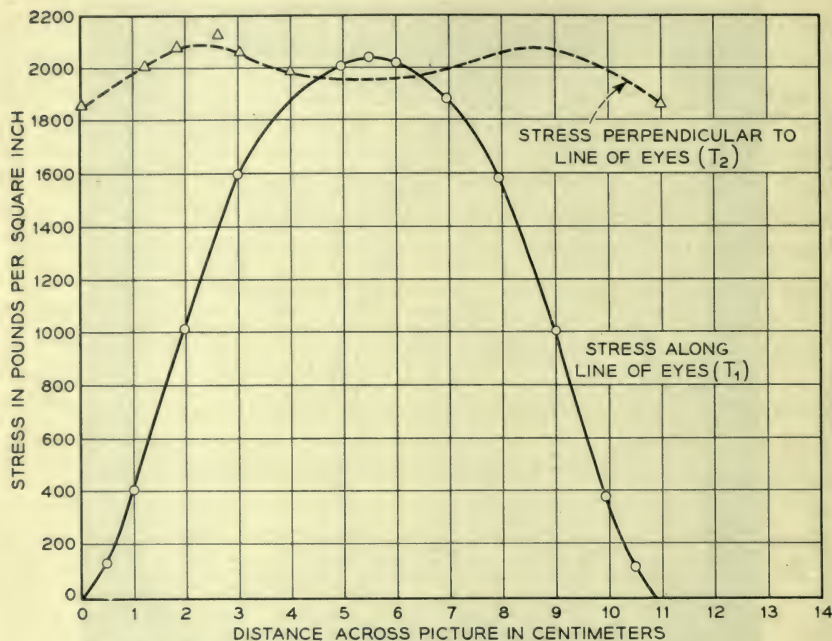


Fig. 5 — Stresses along X and Y directions for a square terminal.

or 71 per cent of the winding stress applied to the wire. The rest of the stress is lost in the contraction that occurs when the wire causes plastic flow to occur at the corners of the specimen. This plastic flow causes the terminal (both bakelite and metal) to flow around the wire and provides an air tight joint which is an essential requirement for a good solderless wrapped connection. That the stress in the terminal is high enough to do this can be seen directly from the photoelastic picture, Fig. 3. Counting the lines as far as they can be distinguished by a microscope, there are 72 lines from the eyes of the picture (which are isotropic points) up to 0.0525 inches from the geometrical corner lines. Using the stress constant of photoelastic bakelite which is 88 pounds per square inch per fringe per inch length along the optic path, this corresponds to a stress per unit area of

$$T = \frac{72 \times 88}{.95} = 6700 \text{ pounds per sq in.} \quad (3)$$

Since the total force put on by the wire is supported by successively smaller cross sections, as one approaches the corners the yield stress of bakelite of 15,000 pounds per square inch will be attained at a radius of

$$\left(\frac{0.0525}{x}\right) \times 6700 = 15,000 \text{ pounds/sq in. or } x = 0.0235 \text{ in.} \quad (4)$$

Hence, plastic flow should occur for about 23 mil inches into the plastic. Unwrapping the wires from the terminal, it is found that depressions of this order are cut in the terminals. Since one of the requirements of a better connection is that an air tight bond shall be formed between the wire and terminal, it is obvious that the terminal should have a low enough yield stress so that a sizable groove can be cut in it by the hoop stress of the wire. This rules out such terminals as hardened steel in the most satisfactory connections. It has been found that copper, brass, aluminum, soft iron and nickel silver are soft enough to meet this requirement. Any material with a plastic flow limit in compression, much lower than photoelastic bakelite, would probably have such a deep groove that it would be difficult to maintain the desired hoop stress.

Using the photoelastic technique as a tool, considerable data has been obtained on desirable shapes for the terminal and limiting winding stresses that can be used. One of the most used terminals is the rectangular terminal and Fig. 6 shows a photoelastic picture of a terminal 0.8 inches by 0.4 inches wound with nineteen turns of 0.050-inch copper wire with a winding load of 28 pounds. This figure is particularly easy to analyze for stress across a line half way down the long edge since the stress along this in the direction of the line varies only a little. Hence,

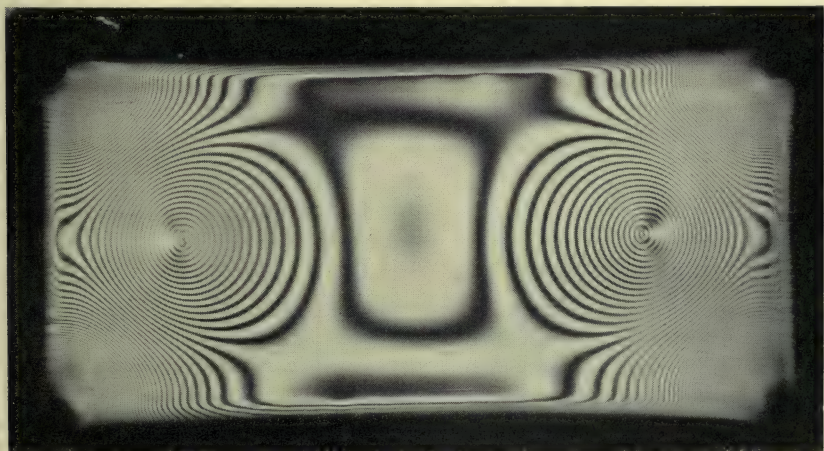


Fig. 6 — Isochromatic lines for a rectangular model 0.4 inches by 0.8 wrapped with nineteen turns of 0.050 inch copper wire with a constant load of 28 pounds.

one can count the number of fringes from the "eye" (which is an isotropic point) to the center of the specimen which in this case is eighteen fringes. Using the stress constant, 88 pounds per square inch per fringe per inch along the optic path, the compressive stress normal to the mid line is

$$\frac{18 \times 88}{0.95} = 1670 \text{ pounds/sq in.} \quad (5)$$

and the hoop residual stress per wire is

$$\frac{1670 \times 0.4 \times 0.95}{2 \times 19} = 16.6 \text{ pounds} \quad (6)$$

which is 60 per cent of the winding stress. Hence, the change in shape has not made any appreciable difference in the residual hoop stress. A specimen 0.4" x 1.6" was also tried and this had a residual stress of 56 per cent of the winding stress.

In order to determine the number of turns required to make a satisfactory joint, measurements were made of the residual stresses as a function of the number of turns with the results shown by Fig. 7. All of these experiments were made with the same weight, 28 pounds, which results in a stress of 14,300 pounds/square inch. Down to five turns, about 50 per cent of the winding stress is maintained. The results are consistent with assuming that the wire unwinds to the extent of two corners on each end while 60 per cent of the winding stress is maintained in all the other turns. As seen from Fig. 8, this is what one might expect, for when the constant tension is released, recovery will cause the

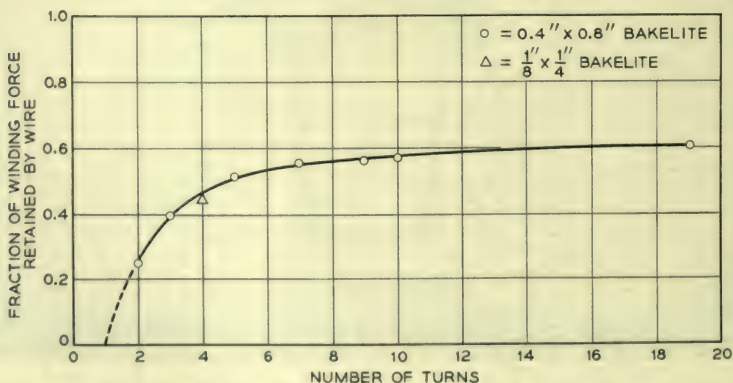


Fig. 7 — Residual hoop stress as a function of number of turns.

first corner to bend out and lose contact with the terminal. When the second corner attempts to unwind, it pulls the first corner up against the sample and no unwinding beyond the second corner can occur. In order to obtain the most satisfactory connections, at least five or six turns should be used.

The fact that the first two corners on each end do not make close contact to the inside terminal produces a very beneficial result when the connection is subject to vibration due to the handling and operation of

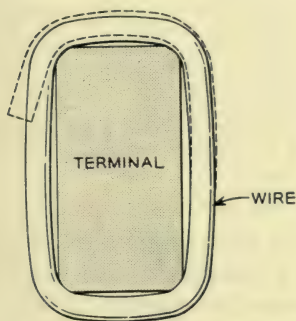


FIG. 8 — Locking in effect in a rectangular terminal which allows unwrapping at only two corners on each end.

a relay. In a soldered connection, large bending strains are caused by vibrations at the point of contact between the wire and the solder and in standard vibration tests when large 60-cycle vibrations are impressed upon the wires, the wires fatigue and break off at the point where the wire enters the solder in times in the order of fifty hours. Similar tests have been carried out for solderless wrapped connections and up to times of 2000 hours and longer no breaks have occurred. This is due to the fact that a bending strain is not enhanced by a sharp discontinuity as it is in the soldered connection and strains for a given vibration amplitude should be less than half as large as those for a soldered connection. Since the relation between fatigue and strain is such that a reduction of strain of two to one or more caused an increase in the number of cycles before breakage of factors of 1000 or more, the increased life under vibration for the solderless wrapped connection is not surprising.

Next a series of measurements were made on the value of the winding stress necessary to preserve a hoop stress in the wire. This was measured by winding wires with different weights on photoelastic samples and measuring the residual hoop stress by the technique described above. This was done for both copper and aluminum wires with the re-

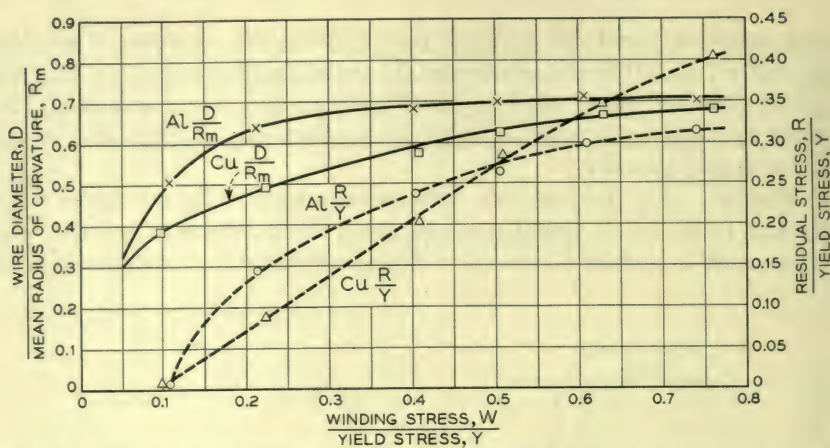


Fig. 9 — Relation between remanent hoop stress, ratio of wire curvature to wire diameter and the applied winding stress for copper and aluminum wire.

sults shown by the dashed lines of Fig. 9. To obtain a hoop stress greater than zero, the ratio of winding stress to yield stress must be greater than 0.1. To obtain a hoop stress that is 0.2 of the yield stress, a winding stress of 35 per cent of the yield stress has to be employed. This is the value recommended to give the most stable terminal. The winding stresses were carried up to 0.8 of the yield stress. At this stress the copper wire at the corners tends to draw down to a low value and may break when it unwinds and hence this is probably the upper limit for winding stresses. The radius of curvature of the middle of the wire, as it is bent around a corner was also measured from photographs similar to that shown on Fig. 6 and the ratio of the wire diameter to the mean radius of curvature is shown by the solid lines of Fig. 9. This is an alternate way of specifying the necessary winding force which may be useful for other shapes of terminals.

While square and rectangular terminals are very satisfactory shapes for the inside terminal, they are not the only ones that can be used. A number of coined and U shaped terminals are in general use as discussed by Mallina. In order to investigate the necessary requirements for such shapes, a number of experiments have been made on circular and elliptical terminals. When a wire under tension is wound around a rod of circular cross section, there are two sets of opposing stresses, one of which tends to make the helix smaller and the other to make it larger. As shown by Fig. 10, the tension in the wire tends to make the helix hug the cylinder while the bending strains introduced by the wrapping of the wire around the cylinder tend to make the helix open up when the constant stress is released. A number of experiments were made on

wrapping copper wire 20 mil inches in diameter on a steel cylinder having a diameter of 0.124 inches. In all cases even up to stresses of 80 per cent of the yield stress, the helix failed to grip the cylinder. Some further measurements were made with smaller sized inner cylinders down to 20 mil inches in diameter. At this small radius the wire barely gripped the cylinder and it took about 70 grams stripping force to pull the wrapped wire off the cylinder. As seen from Fig. 11, the normal force per unit length against the cylinder is balanced by the hoop force in the wire according to the equation

$$\theta F_N r = 2F_H \sin \frac{\theta}{2} \quad \text{or} \quad rF_N \doteq F_H. \quad (7)$$

The stripping force SF , when the wire does not dig into the terminal, should be equal to

$$SF = 2\pi r n F_N f = 2\pi n f F_H, \quad (8)$$

where n is the number of turns, f the coefficient of friction which is about 0.15 to 0.2 between metals. For a stripping force of 70 grams for 6 turns, the remaining hoop stress is equal to

$$F_H = 9.3 \text{ grams.} \quad (9)$$

Since the wire was wound with a 700-gram force, it is evident that only about 1 per cent is maintained in the wire, which is entirely inadequate.

In order to obtain a good wrapped connection with high hoop stress in the wires, some means has to be employed to eliminate the unwraping effect of the strain due to bending. This can be accomplished by changing the shape of the terminal from a circular cylinder to a dissymmetrical shape. For then, as shown by Fig. 8, the tendency to un-

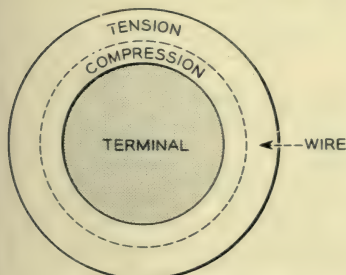


FIG. 10 — Strains in a wire wrapped around a circular terminal.

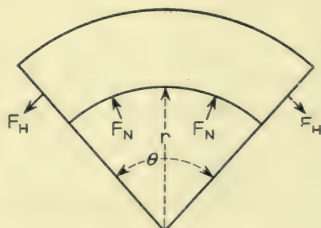


Fig. 11 — Relation between hoop stress and stripping force for terminal not indented.

wind is opposed by the locking in effect of the dissymmetry which in the most preferred types of terminals, as in Fig. 8, comprises abrupt changes in direction around the periphery of the terminal cross section. Some experiments were made with a cylinder whose cross section, as shown by Fig. 12, consisted of parallel sides with a semi-circle on each end. When the length A was twice the width, the top curve of Fig. 12 shows the stripping force as a function of the winding force. Assuming that most of the grip occurs on the two semi-circular surfaces, the hoop

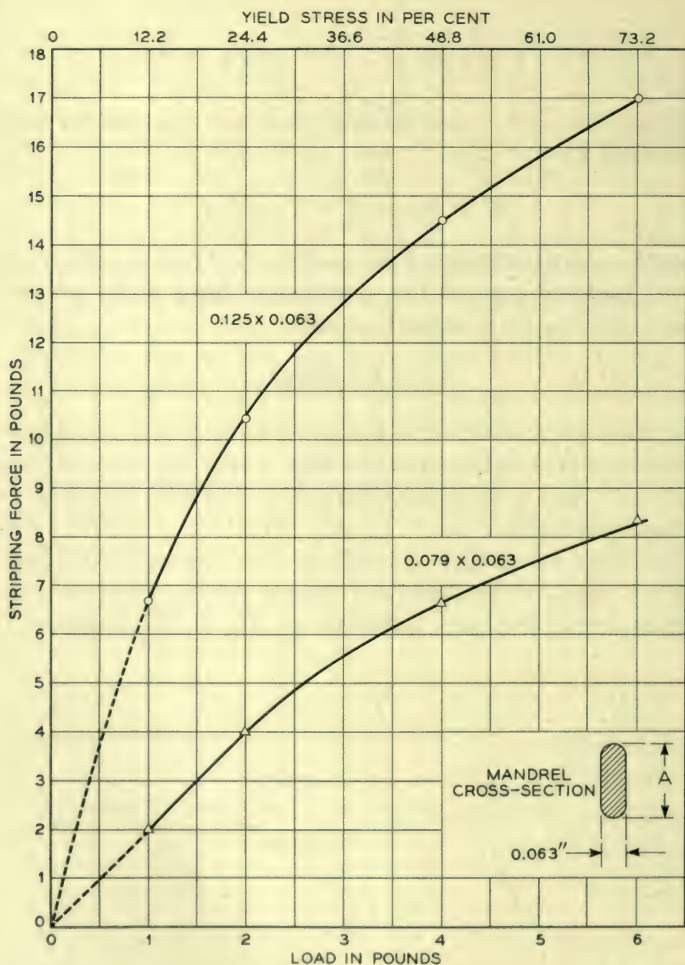


Fig. 12 — Stripping force as a function of winding load for an elliptical type terminal.

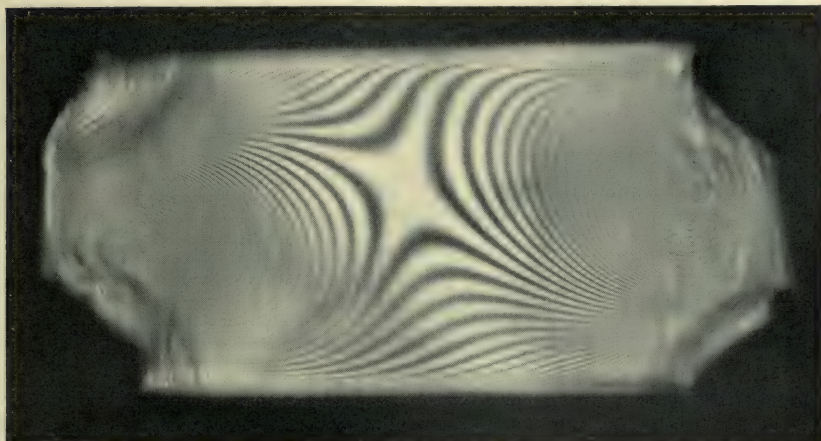


Fig. 13 — Photoelastic picture of a terminal when the outside wire diameter approaches the terminal diameter.

stresses indicated are in the order of 50 to 60 per cent of the winding stresses in agreement with the photoelastic experiments. Ratios of 5 to 1 on the length-width ratio were also tried with substantially the same result. To determine how much dissymmetry is necessary to lock in the bending stresses, a ratio of 1.25 to 1 of the length to width was tried with the result shown by the lower curve of Fig. 12. The stripping force for this ratio is about half that for the larger ratios. The conclusion is that the stripping force decreases as the symmetry increases but that a small deviation from circular symmetry is sufficient to lock in the bending stress. However, for the more satisfactory connections, other factors such as adequate intimate gas tight contact areas indicate that the required dissymmetry involves abrupt surface changes in the nature of edges having appreciable penetrating power with respect to the wire. Therefore, while a terminal like that of Fig. 12 may have the ability to lock in the bending stresses, it may not, for other reasons, be the better terminal to use.

All of the photoelastic and other types of stress measurements were made for terminals that have a large stiffness in torsion and as seen by the photoelastic pictures, the stresses are nearly plane stresses, i.e., they are all tensions, compressions or shears in a plane perpendicular to the axis of the connection. If, however, the torsional stiffness becomes low, another type of deformation can take place, namely, a twist of the whole terminal due to the torque put on by the helical form of the winding. Fig. 13 shows a photoelastic picture of the strain in the terminal when

the size of the outside wire is comparable to the size of the terminal and it is obvious that a twist in the terminal is occurring. This is a case of three dimensional stress rather than plane stress and cannot easily be analyzed from the photograph. The photograph does, however, show a twist of the terminal.

The twisting strain can be most easily analyzed by taking a long section of terminal of low torsional stiffness, winding 100 or more turns on the terminal and measuring the angle of twist as discussed in the paper by Mallina. Calculations by Love¹ show that a twist in an elliptical section with its length along the Z axis introduces shearing strains in the X, Z and Y, Z planes, i.e., $e_{zx} = S_5$ and $e_{yz} = S_4$ shearing strains equal to

$$S_5 = -\tau \left(\frac{2a^2}{a^2 + b^2} \right) y; \quad S_4 = \tau \left(\frac{2b^2}{a^2 + b^2} \right) x, \quad (10)$$

where $2a$ is the diameter of the ellipse along the X direction and $2b$ the diameter of the ellipse along the Y direction and τ the angle of twist in radians per centimeter. For the terminal with 100 turns of 0.020 mil copper wire discussed by Mallina whose data are given by Fig. 13, 45° angle of twist occurs in 2 inches giving a value of $\tau = 0.157$. This causes a shearing strain of about 2 per cent in the worst case which is enough to cause a considerable permanent set. While this twist is useful in studying stress relaxation in the wire, it is undesirable for a solderless wrapped connection to have too much twist since it may cause the terminal to twist off in the winding process. According to the data of Fig. 13 of Mallina's paper, no terminal set occurs for nickel silver if the twist in radians per centimeter is less than 0.09 which corresponds to a maximum shearing strain in the X, Z plane of 1.1 per cent. Hence in order to avoid excessive permanent set and twisting off of the inner terminal, the size and shape of the terminal should be controlled so that shearing strains due to twisting should be less than 1 per cent. For standard shapes such as rectangles and ellipses formulae are available to relate the maximum strain to the dimensions of the terminals and the moment due to the winding stress. For a given wrapping tension, this moment can be calculated from Appendix I of Mallina's paper.

Summarizing the results of this section, the necessary conditions that the terminal should meet are:

1. The wrapped connection is held together by the hoop stress in the outside wrapping wire. This can be locked in if the terminal has a dissymmetrical shape in which the length-width ratio is 1.5 or greater,

¹ Love, Theory of Elasticity. Chap. XIV, p. 310, 4th Edition, Cambridge University Press.

or if some regular shape such as a square, rectangle or rhombus is used. Sharp corners are helpful since a sharp bend in the wire occurs around them.

2. The material of the terminal must be strong enough so that the wire will not deform or cut through the terminal but must be plastic enough so that an appreciable groove can be cut in it by the hoop stress of the wire, in order that an air tight connection shall be made. The most satisfactory metals are brass, copper, soft iron, nickel silver and aluminum.

3. In the most satisfactory solderless wrapped connection, the wrapping wire unwinds to the extent of half a turn on each end and about six turns or more are required to make a good connection.

4. To maintain sufficient hoop stress, the constant wrapping stress should be from 0.2 to 0.7 of the breaking stress of the wire.

5. Shearing strains in planes parallel to the axis of the terminal should not exceed 1 per cent in order to eliminate terminal set.

PHOTOPLASTIC ANALYSIS OF STRAINS IN WIRES OF A WRAPPED SOLDERLESS CONNECTION

All of the strains in the inner block or terminal are elastic except at the corners. Hence, for the interior terminal, photoelastic bakelite is a satisfactory material for strain investigations. However, the outer wire is necessarily stressed beyond its elastic limit and ordinary photoelastic techniques cannot be applied. In order to see if the wire strains could be studied with photoelastic bakelite, some time was spent in heating rods to a temperature for which they become elastic, winding them under a stress and cooling under the applied stress. Although the winding process was carried out successfully several times, the bakelite rod always broke on cooling. This appeared to be due to the fact that bakelite is nearly linear up to the breaking point, i.e., it suffers from brittle fracture and does not simulate a metal in this respect.

Some measurements had previously been made at the Bell Laboratories and in England² on polyethylene which indicated that it had properties similar to a metal in the plastic range. Stress-strain curves up to 15 per cent strain are shown by Fig. 14, and it is evident that on the ascending part, the curve is very similar to that for copper or soft iron. On the relief from stress, however, a considerably larger recovery is obtained than for a metal. This material is fairly transparent and the lower curve marked R/t shows the relative retardation for a 5461 Å°

² Miss S. M. Crawford and Dr. H. Kolsky, Stress Birefringence in Polyethylene. Proc. Phys. Soc., Section B, London, **6**, Part 2, pp. 119-125, Feb. 1, 1951.

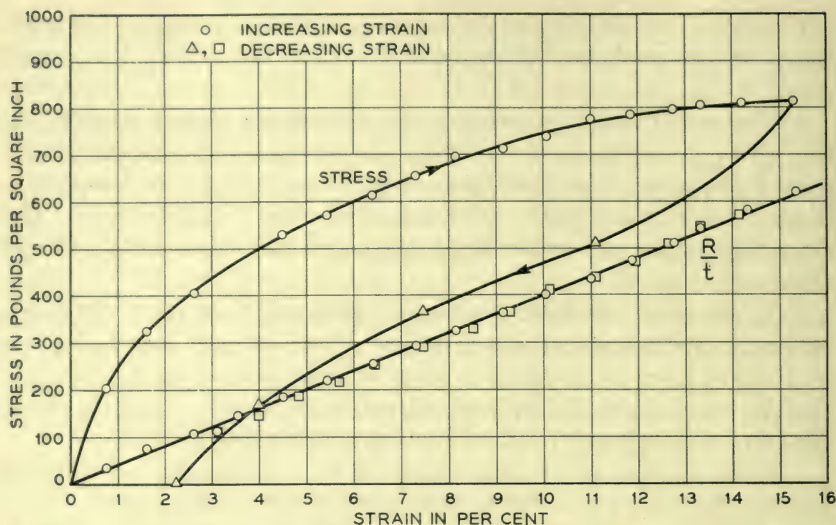


Fig. 14 — Stress and birefringence strain characteristic of quenched polyethylene.

mercury line. This relative retardation R/t is related to the number of fringes N and the wavelength λ by the equation

$$\frac{R}{t} = \frac{N\lambda}{t} \quad (11)$$

and hence for a 1 per cent strain there are 7.3 fringes for a 1-cm path length in the optic direction. It will be observed that for both increasing and decreasing strains, the birefringence is directly proportional to the strain. Hence, by using polyethylene it appeared possible to measure the strain even in the plastic region.

The first experiment tried was to wrap a square metal rod with a polyethylene "wire" one-sixteenth inch in diameter with a wrapping stress about half the yield stress. It was found, however, that the wire sprang off the metal rod when the constant stress was released. This is due to the fact that the polyethylene has considerably more recovery than the metal wire and the dissymmetry is not sufficient to lock in the bending stress. This shows that one of the requirements of the wire is that the recovery shall not be too large.

If we are to use polyethylene as a photoplastic material, it is necessary to simulate the unloading curve as well as the loading curve. This can be done by heating up the polyethylene when it is wound under a

load, cooling under a load, and then removing the load at room temperature. Fig. 15 shows the loading and unloading curves for polyethylene as a function of temperature. The stress required to produce a given strain decreases very rapidly as the temperature increases, although the recovery remains about the same irrespective of the temperature. Suppose now that we apply a load and cool the polyethylene down to room temperature maintaining the strain. When the weight is taken off, the unloading curve will parallel that of the 20°C curve, and 42 per cent recovery will be obtained at 50°C and 12 per cent for 90°C.

In order to see if a wrapped joint would be simulated by this means, a one-quarter inch rod of polyethylene was heated up to 97°C, was wound with a one pound winding weight and was cooled to room temperature with the weight attached. This was sufficient to prevent the "wire" from unwrapping and when the weight was removed, the polyethylene "wire" gripped the metal closely and formed a bond similar to the wrapped connection. In order to analyze the strain, one has to cut a section through the center of the wire and put the section in the polariscope. First, a number of unstrained polyethylene samples were cut by various techniques and it was found that if they were cut with a

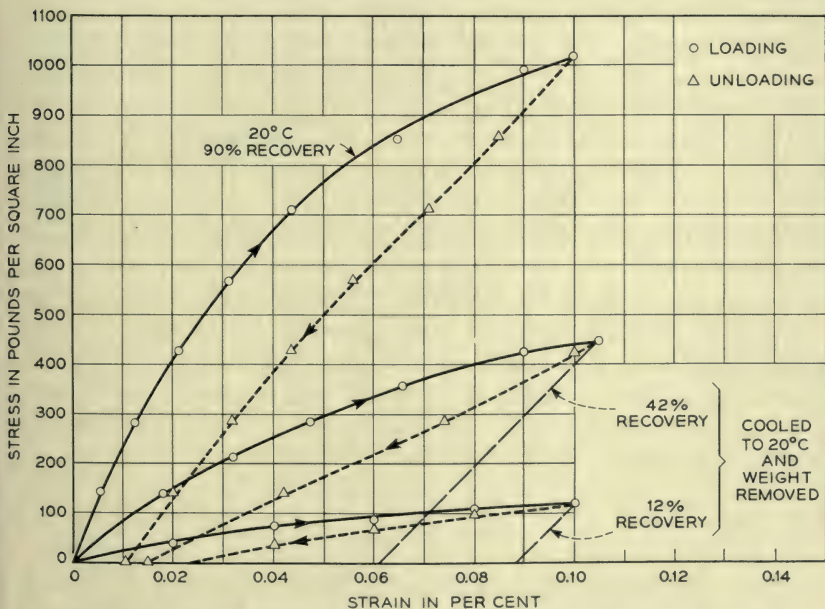


Fig. 15 — Stress strain recovery curves for quenched polyethylene.

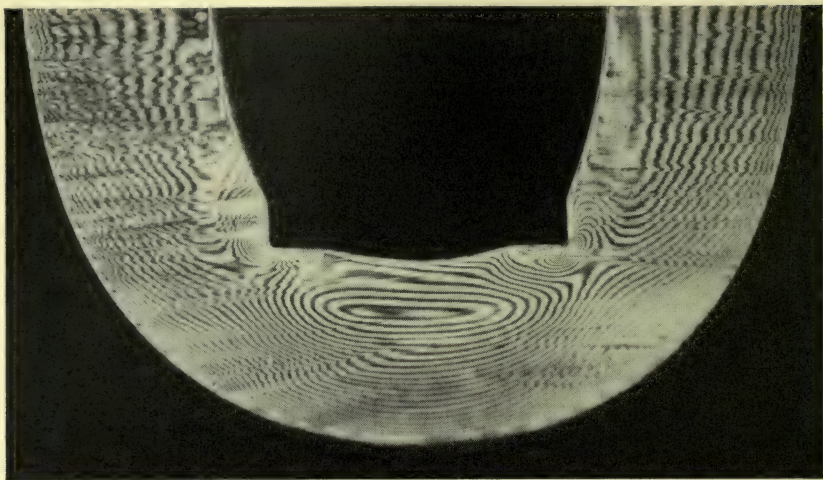


Fig. 16 — Photoelastic picture of a polyethylene "wire" wrapped at 97°C around a rectangular terminal and cooled off under the applied load.

jewelers saw with a good deal of set to the teeth, no strains were introduced in the sawing process. Using a jig with two parallel guides a sample 0.040 inches thick was cut through the polyethylene wire. Taking one in the center of the wrap, a photoelastic picture was taken with the result shown by Fig. 16 with an enlargement of one corner shown by Fig. 17.

The easiest parts to analyze are the strains in the two legs of the sample since the strains are similar to those for a bent section. The long leg which was twice as long as the short leg has its zero order fringe along the inside edge. There are twelve lines to the outer edge which corresponds to a tensile strain of 16 per cent with an average tensile strain of about 8 per cent which is about the strain caused by the load-ing weight. In the short leg the zero order fringe is the inside oval and the strain is about 11 per cent compressive at the inner edge of the segment and about 38 per cent tensile at the outer edge. These values are consistent with the radius of curvature that the wire is bent around for it can be shown that the strain in a wire of diameter d bent around a cylinder of diameter D without tension is equal to

$$S = \frac{2\rho}{D + d}, \quad (12)$$

where ρ is the radial distance measured from the center of the wire outward, D the diameter of the cylinder and d the diameter of the wire. If ρ is positive or the point is outside the center line, the strain is posi-

tive or tensile, while if ρ is inside the center line the strain is negative or compressive. From measurement of Fig. 16, it appears that the equivalent inner cylinder that corresponds to the radius of the center of the wire is about 3.0 times the wire diameter and hence the strain should be 25 per cent tensile at the outer edge and 25 per cent compressive at the inner edge. The addition of a tension of 13 per cent due to the winding force makes the outer strain 38 per cent and the inner one 12 per cent compressive which are close to the values found. This indicates that the tensile strain is somewhat higher in the short leg than in the long one. The same tensile strain occurs around the outside periphery of the wire opposite the corners of the terminal, but a very high

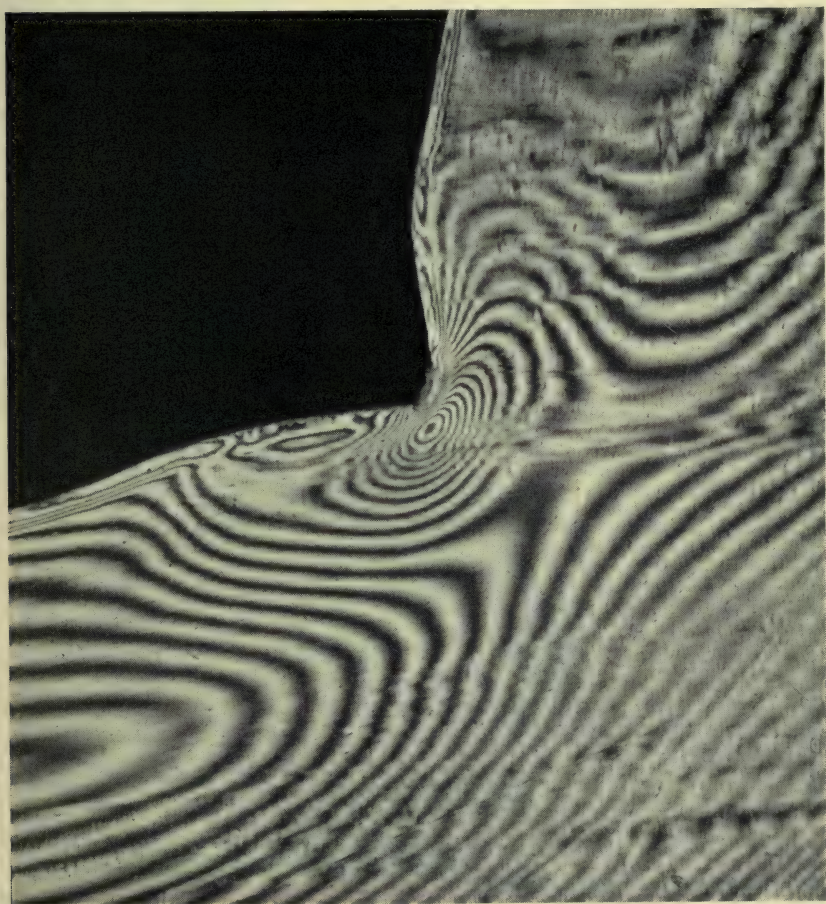


Fig. 17 — Enlargement of one corner of the photograph of Fig. 16.

point of compression develops just below the point of contact between the wire and terminal. The distribution of compressive and tensile strains in the wire is shown pictorially by Fig. 18. The distribution of strains in a metal wire can be considered to be quite similar except that the tensile strain due to winding should be only 2 to 3 per cent as seen from Fig. 4, while the strains due to bending may be even higher, for as seen from Fig. 9 the ratio of D/d may be in the order of 1 and the bending strains may be as high as 50 per cent.

In order to specify the properties that the wire must have in making a good solderless connection, experiments have been made on how much recovery can be tolerated in the wire. Since it is difficult to get a series of metal wires having different amounts of recovery, the technique was resorted to of heating polyethylene to a definite temperature, winding under a load equal to half the yield stress at the winding temperature, and cooling to room temperature under the load. As shown by Fig. 15, known recoveries can be obtained in this way. It was found that the largest amount of recovery that could be tolerated to make a joint at all was 20 per cent while a reasonable hoop stress was not obtained until the recovery was less than 10 per cent.

In summary, the necessary conditions that the wire should fulfill are:

1. Since strains of 50 per cent may be encountered in bending wires around sharp corners, wires should be used which have a large difference between the yield strain and the breaking strain. Copper, aluminum and soft iron are materials of this class while phosphor bronze and music wire are not as satisfactory.

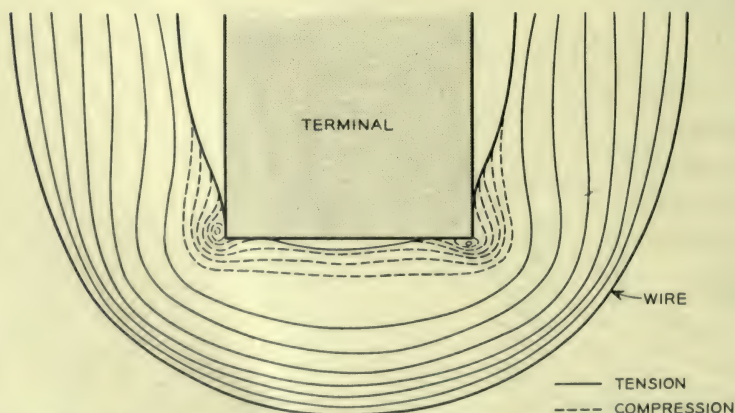


Fig. 18 — Distribution of compressional and tensile stress in the wire of a wrapped solderless connection.

2. The amount of recovery from strains of 20 to 50 per cent should not be greater than 20 per cent.

PERMANENCE OF WRAPPED SOLDERLESS CONNECTIONS

By the photoelastic, photoplastic and strain analysis of the previous two sections, it has been demonstrated that the wrapped solderless connection is held together by the hoop stress in the outside wire whose value is determined by the winding stress and the locking in effect dependent on a dissymmetry of the terminal. The high stresses cause plastic flow in the wire and terminal in such a manner that the two materials flow together and produce an intimate air tight joint. The intimate nature³ of this contact has been demonstrated by dip coating nickel silver terminals with pure tin and wrapping them with cleaned bare copper wire. The wrapped terminals were then placed in a glass tube, evacuated, sealed off and heated for 400 hours to 180°C (37°C below the melting point of tin). The samples were then removed, mounted vertically, polished, etched and examined microscopically for distinguishing constituents. It is believed that if such a constituent appeared on the originally bare copper wire after such treatment, that the contact was sufficiently intimate to permit solid state diffusion. A section of the wire in contact with the corner is shown by Fig. 19. The copper is seen

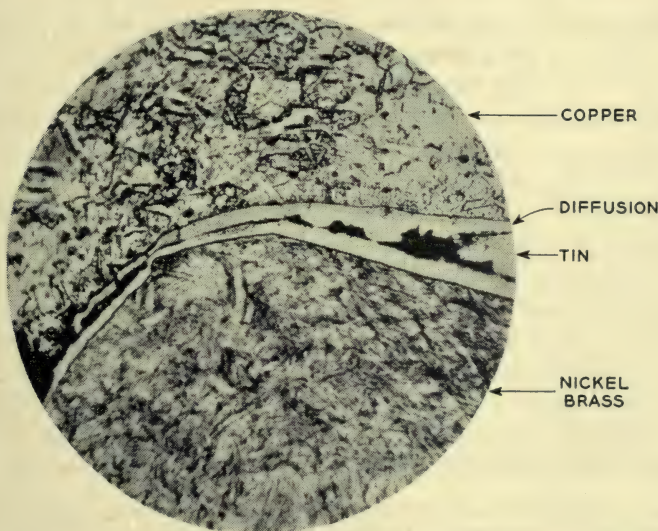


Fig. 19 — Solid state diffusion of tin into copper in a wrapped solderless connection.

³ This experiment was conducted by G. S. Phipps.

to have a heavy layer of tin constituent at the contact surface. From the test, it can be concluded that wrapped connections are sufficiently tight to allow solid state diffusion and are therefore good electrical contacts.

On the other hand the contacts are not welded contacts such as occur when two pieces of aluminum are cold pressed together with strains in excess of 75 per cent. This is shown experimentally by the simple process of unwinding the wire from the terminal which takes place with no excess force when the wire is removed from a terminal corner. Since the strains at the points of contact do not exceed 30 to 40 per cent, one would not expect cold welding. It is possible, however, that in tinned terminals, some long time diffusion takes place at room temperature in the manner demonstrated at higher temperatures by the data of Fig. 19. This would occur very slowly and cannot be relied upon solely to maintain the contact.

Hence, it appears that long life in the connection depends on maintaining sufficient hoop stress in the wrapping wire to keep the elements of the connection sufficiently tightly pressed together so that no corrosion can occur in the connection in such a manner as to interrupt the electrical continuity. This is a problem in stress relaxation rather than creep. Stress relaxation is intrinsically a simpler phenomenon since the major fraction of stress that can be relaxed will be relieved through viscous flow in previously formed slip bands or along grain boundaries, and no generation of new slip bands is required. However, under ordinary creep conditions, an increase in stress is presumably attended by the generation of new slip bands. It appears likely then that stress relaxation phenomena even at quite high stresses should more nearly follow the conditions that have been established for low stresses than would be the case for creep phenomena.

A good deal of work has been done on stress relaxation at low stresses, particularly by Zener⁴ and his coworkers, and this will be briefly reviewed. According to these studies, stress relaxation can be caused by several mechanisms including stress induced migration of impurities in the metal, viscous behavior of slip band material and the viscous behavior of grain boundaries. At the common junction between the two metal grains, there is an amorphous layer of material which acts as a viscous medium, i.e., if there is a shearing stress applied across it, the two grains will move with respect to each other with a velocity

⁴ Zener, C., *Elasticity and Anelasticity of Metals*. University of Chicago Press, 1948. T'ing-Su Ké, *Experimental Evidence of the Viscous Behavior of Grain Boundaries in Metals*. *Phys. Rev.*, **71**, No. 8, pp. 533-546, April 15, 1947. T'ing-Su Ké, *Anelastic Properties of Iron*. Tech. Publication No. 2370, Metals Technology, June, 1948.

$$V = \eta T/D, \quad (13)$$

where D is the thickness of the layer, η the coefficient of viscosity and T the shearing stress. Hence no matter how small the shearing stress, one grain will move with respect to the other in a finite time. The amount that the grains can move is limited by the necessity of making the grain boundaries fit. According to Zener, the situation is analogous to the case of a jigsaw puzzle in which the overall configuration possesses rigidity in spite of the fact that no shearing stress exists between adjacent pieces. Zener has calculated that the ratio of the relaxed stress to

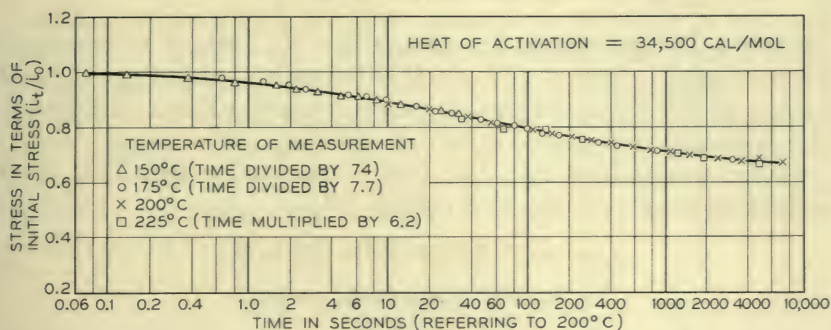


Fig. 20 — Stress relaxation in aluminum at three temperatures for strains less than 10^{-4} (after Ké).

the initial stress is equal to

$$\frac{T_R}{T_I} = \frac{\frac{1}{2}(7 + 5\sigma)}{7 + \sigma - 5\sigma^2}, \quad (14)$$

where σ is the value of Poisson's ratio. For values of Poisson ratio of from 0.25 to 0.5 this ratio lies between 0.595 and 0.76.

Fig. 20 shows measurements of stress relaxation plotted against time for aluminum for three different temperatures. These were obtained⁴ by twisting an aluminum wire through a definite angle and observing the force required to hold it at this angle as a function of time and the temperature of the wire. All the curves can be made to coincide by multiplying the times by different factors. If we define the relaxation time τ as the time required to relax half of the variable component of stress, i.e. $\frac{1}{2}(1 - 0.67)$ of the stress, this relaxation time fits an equation of the form

$$\tau = Ke^{H/RT}, \quad (15)$$

where K is a constant, H an activation energy, T the absolute tempera-

ture in degrees Kelvin and R the Boltzman constant for one gram mole of the material. R is closely equal to 2 calories per degree K. Hence, if H is expressed in calories per gram mole and the value of K is obtained to fit equation (15), we have

$$\tau = 9.2 \times 10^{-15} \times e^{\frac{34,500}{2T}}.$$

The constant K is close to that given by the Langmuir-Dushman⁵ theory

$$K = \frac{hN}{H} = \frac{6.62 \times 10^{-27} \times 6.06 \times 10^{23}}{34,500 \times 4.187 \times 10^7} = 2.7 \times 10^{-15}, \quad (16)$$

where h = Planck's constant equal to 6.62×10^{-27} ergs, N is Avogadro's number equal to 6.06×10^{23} and H is the activation energy expressed in ergs. Su Ké⁴ has shown that the activation energy for grain boundary slip is essentially the same as for self diffusion and for creep.

Similar results have been found for α -brass and α -iron. These have activation energies shown by equation (17)

$$\begin{aligned} \alpha\text{-brass} & \text{ 41 kilocalories per mole,} \\ \alpha\text{-iron} & \text{ 85 kilocalories per mole.} \end{aligned} \quad (17)$$

Although measurements have not been made for copper, the activation energy of self diffusion is about 57.2 kilocalories⁶ per mole, but 39.9 kilocalories for the principle impurity silicon.

All of these measurements were made for strains under 10^{-4} , and the question arises as to whether these concepts are valid for the much higher strains experienced in the wrapped solderless joint. From the photoelastic pictures, Figs. 16 and 17, it is obvious that the greatest stress inhomogeneity occurs in the neighborhood of the corners and flow will take place in such a way as to relieve the high stress concentration. This will have the effect of making the terminal and wire mate even closer and may result in a slight transient lowering of the hoop stress. After the initial formation, however, it will be the long time relaxation of the hoop stress in the wire that determines the lasting quality of the joint.

As discussed previously, the twist that the terminal takes is determined by the mean value of the hoop stress in the wire, and any relaxation in this hoop stress can be studied by observing the angle of twist as a function of time and temperature. By using a long terminal wound

⁵ S. Dushman and I. Langmuir, Phys. Rev. **20**, (1922) p. 113, 1922.

⁶ Zenner, Elasticity and Anelasticity of Metals. Table 12, p. 98, Chicago University Press.

with 100 turns or more of copper wire, a twist of 50° or more can be obtained which is sufficient to measure.

In order to test first the relaxation in the copper wire alone, the inner terminal was made of clock spring steel 0.0124 inches by 0.062 inches. This was wound with 100 turns of 0.020 inch copper wire tensioned at 2.87 pounds (9,300 pounds per square inch). A twist of 25° was obtained which is sufficiently large to measure. If one observed the angle after transient creep has occurred, the angle decreased on the average about 17 per cent in the first month as shown by the circles of Fig. 22. The values agree with the solid curve which was established by relaxation measurements as a function of temperature as discussed in the next paragraph. At room temperature, further decreases in the angle of about 10 per cent were observed out to times in the order of a year.

If, however, the wrapped connection was heated up, a faster relaxation of the hoop stress occurred. Fig. 21 shows the ratio of angle measured to the initial angle as a function of time when the connection is subjected to a temperature of 200°C . As shown by the dashed line, which is a plot of the exponential equation

$$R = e^{-\alpha t} \text{ where } \alpha = 2 \times 10^{-4}, \quad (18)$$

this is not a single relaxation of the type found for grain boundary motion but is a sum of effects occurring with different activation ener-

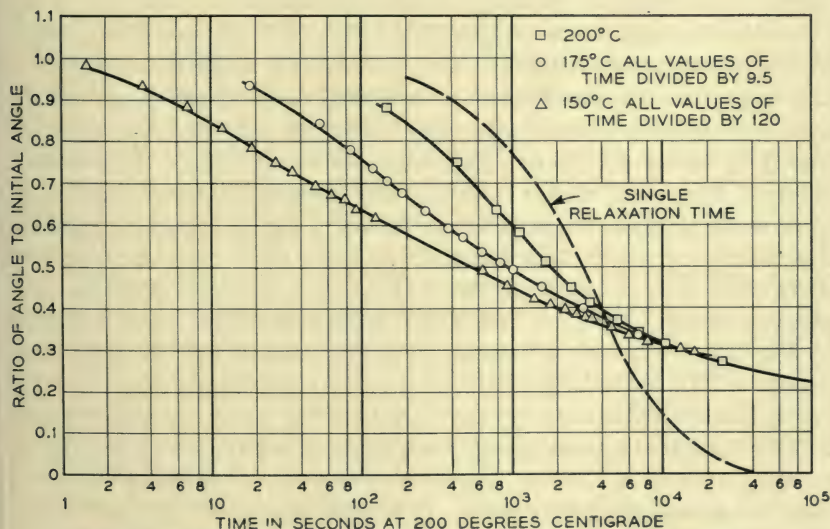


Fig. 21 — Ratio of angles of twist at time shown to initial angles.

gies. Furthermore, the change is not limited to two-thirds as in the case of grain boundary motion, but is much more tending toward a value of 0.2 for very long times. It appears that several processes are involved in addition to grain boundary motion. These are probably connected with slips along slip planes which occur with a lower activation energy when the stress is high. As the slip increases, strain hardening occurs with a resultant increase in activation energy until the activation energy of grain boundary movements is reached. For 0.3 relaxation and lower, the activation energy remains constant and equal approximately to 40 kilocalories per mole — the self-diffusion value — as can be calculated from the 150°C, 175°C and 200°C relaxation curves of Fig. 21. These curves show that the activation energy varies from 12.5 kilocalories at 0.9 relaxation, to 40 kilocalories for long time effects.

Similar measurements have been made on nickel silver terminals which are the terminals actually used and as seen from Fig. 21 of Mallina's paper, these agree quite well with those measured for spring steel terminals. The nickel silver terminals had the dimensions 0.0148 inches by 0.062 inches. A twist of 46° was obtained for 100 turns of 0.020 inch copper wire with 2.87 pounds winding stress applied. At this angle of twist, a permanent set of 19° occurred when the outside wire was unwound. If we subtract that value from the initial twist, the time-angle curve is very similar to that for spring steel and indicates that no additional relaxation occurs in the nickel silver terminal.

The conclusion from these experiments is that stress relaxation for the value of strain used in the wrapped solderless connections follows a similar activation energy pattern to that followed for smaller strains except that instead of a single process with a single activation energy we are dealing with many separate processes having an activation energy range from 12.5 kilocalories to 40 kilocalories. For each stage of the process a different activation energy is effective. For example, for a ratio of relaxed stress to initial stress of 0.9 the curves of Fig. 21 indicate an activation energy of about 12.5 kilocalories. With this value of activation energy the time required to relax this amount of stress at room temperature of 25°C (77°F) is 2.98×10^5 seconds or 0.0095 years as shown by Fig. 22. For a ratio of relaxed to initial stress of 0.8, the activation energy is 15.3 kilocalories and the time at room temperature is 0.126 years. Similar values can be calculated for the other relaxation ratios and the complete relaxation ratio and time curves are shown by Fig. 22 for temperatures of 77°F and 135°F. To reach a value of 0.5 of the initial hoop stress requires 2500 years at 77°F and about forty years at 135°F. The circles show the measured values at 77°F carried out in a tempera-

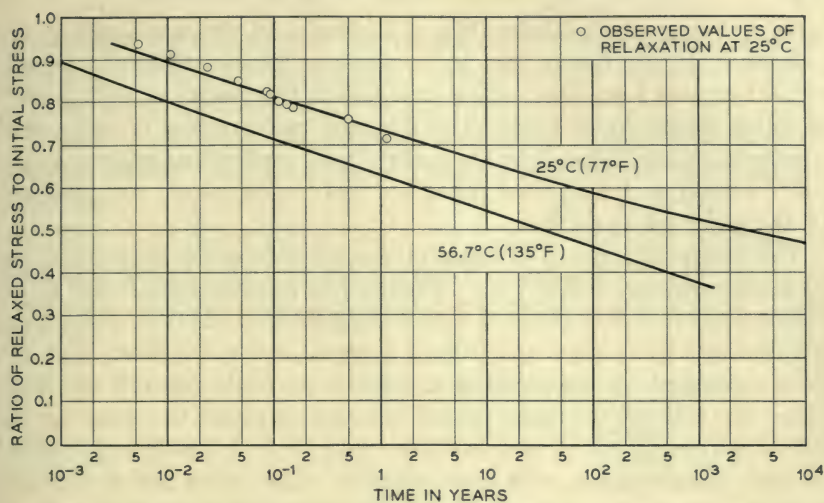


Fig. 22 — Aging at room temperature and at 135°F plotted as ratio of angle to initial angle as a function of time.

ture controlled air conditioned room and these agree well with the calculated values. 135°F is in general the maximum temperature that wrapped solderless connections will be subjected to. For cases of very high temperature it is planned to use copper covered soft iron wires for the wrapping wires since, as shown by Equation (17), iron has a much higher activation energy than copper and can be expected to maintain its hoop stress for forty years even under an ambient temperature of 200°C.

The question arises as to whether the hoop stress of a small part of the initial value, that has been shown to continue for a long time by the data of Fig. 21, is sufficient to maintain a good contact. This question may be important if aluminum is to be used as a wrapping wire since with an activation energy of only 34.5 kilocalories, 0.5 of the hoop stress will be maintained for forty years only for temperatures lower than 100°F. The problem then is whether corrosion can occur between the wrapping wire and the inside terminal when the relaxable component of hoop stress has been relaxed. A very sensitive test for this question is obtained by winding aluminum wire on an aluminum terminal since if any break occurs in the contact between the wire and the terminal, oxidation of the aluminum surface takes place very rapidly and should affect the resistance of the solderless connection. Accordingly, a number of aluminum-aluminum solderless connections were wound up and their resistances were measured. They were then put in an oven and heated to a temperature of 200°C for twelve hours, which was suffi-

cient to relax all the stress that can be relaxed. On remeasuring the resistance, it was found that there was no change within the experimental error of 1 per cent, which corresponds to a resistance of 3×10^{-5} ohms. A similar result is found by studying the corrosion of surfaces of copper and nickel silver in solderless wrapped connections when they are fully relaxed and subjected to a corrosive atmosphere as discussed in the paper by Van Horn.

The stripping force of the aluminum-aluminum connection subjected to a temperature of 200°C for 12 hours has actually been found to increase by a factor of about 2 which suggests that the two parts have diffused into each other and formed a permanent connection. This has been confirmed by the increase in force required to unwrap the wire. Since the activation energy of self diffusion is about the same as the activation energy for stress relaxation, then as the hoop stress is relaxed at high temperatures, solid state diffusion takes place and a diffusion joint is formed in aluminum. The same process, both for stress relaxation and diffusion, should take place at a much lower rate at lower temperatures and as the hoop stress is relaxed, a diffusion connection between the two parts is formed so that no decrease in the conductivity of the connection occurs and an actual increase in the strength of the connection results. The same process should result between any two materials provided the energy of self diffusion from one into the other is less or equal to the activation energy of stress relaxation for the weakest component.

It is planned to tin plate both terminals and wires for all of the wrapped solderless connections used in the telephone system. In order to find how the two processes of stress relaxation and self diffusion, which progress as a function of time and temperature, affect the two fundamental properties electrical conductivity and mechanical strength of the connections, a large number of connections were wound up and subjected to temperatures of 200°C for times corresponding to 0.9, 0.8 etc., of the initial stress as determined from Fig. 21. The electrical resistances of the connections were determined before and after the treatment and the stripping force was also measured. Within the experimental error the resistances of the connections remained the same while the average stripping force for twenty connections for each point are shown by Fig. 23. No significant change in the stripping force occurred out to values of stress relaxation less than 0.2 times the initial hoop stress. These experiments show that as the hoop stress is relaxed by time and temperature, self diffusion occurs between the two parts of the connec-

tion in such a way that the mechanical strength and conductivity are maintained unchanged with time.

In summary it has been shown that

1. The connections are sufficiently intimate to permit solid state diffusion but the strains are not high enough to cause cold welding of the connection.

2. The hoop stress relaxes as a function of time and temperature according to well known activation energy equations with an activation energy for copper wires on nickel silver connections varying from 12.5 to

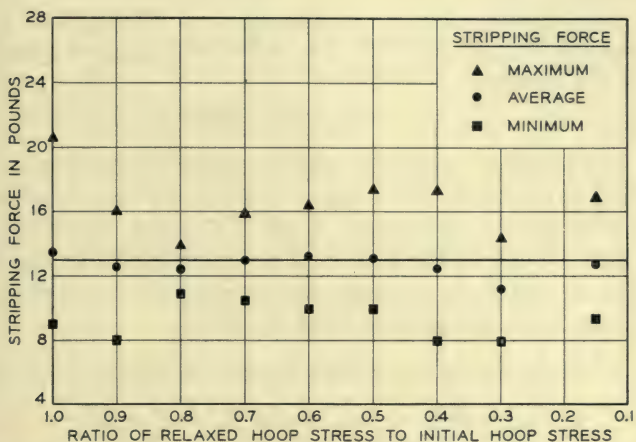


Fig. 23 — Stripping force plotted as a function of stress relaxation.

40 kilocalories. It requires about 2500 years to relax half the hoop stress at 77°F (25°C) and 40 years at 135°F.

3. The twin processes of stress relaxation and self diffusion occur in such a way as to maintain or increase the strength of the connection and to leave the resistivity of the connection unchanged with time.

4. The conclusion from all of these tests is that the life span of the solderless wrapped connection appears to be ample for meeting any of the likely requirements. This conclusion is reinforced by life tests of a large number of wrapped solderless connections in service that have been carried out over two years without a failure and by the long and satisfactory tests of the screw connection whose success also depends on the stress relaxation, diffusion processes. Such connections have been shown to be satisfactory over periods of time in excess of twenty years.

APPENDIX

METHOD FOR ANALYZING STRESSES AND STRAINS FROM PHOTOELASTIC PICTURES

The methods for analyzing photoelastic pictures are given in detail in a number of books and other publications⁷ and only a brief summary of the method used here will be given.

When polarized light is sent normal to a plane of photoelastic material that is strained in the plane, the light is broken up into an ordinary and an extraordinary ray that travel with different velocities. It is shown⁷ that the birefringence, which is defined as the difference between the two indices of refraction μ_1 and μ_2 (i.e. the index of refraction is the ratio between the velocity of light for one of the rays in a vacuum to the velocity in the medium) is given by

$$B = \mu_1 - \mu_2 = C' \sqrt{(T_1 - T_2)^2 + 4T_6^2}, \quad (19)$$

where C' is a constant called the relative stress optical constant, T_1 is the tensional stress along the X axis, T_2 the tensional stress along the Y axis and T_6 the shearing stress in the XY plane. If we change the direction that we call the X axis until we reach the direction of maximum stress, the relations between this stress and the tensional stress at right angles to it are given by the equations

$$\begin{aligned} T'_1 &= T_1 \cos^2 \theta + 2 \sin \theta \cos \theta T_6 + T_2 \sin^2 \theta, \\ T'_2 &= T_1 \sin^2 \theta - 2 \sin \theta \cos \theta T_6 + T_2 \cos^2 \theta, \end{aligned} \quad (20)$$

where

θ is the angle between the X axis and the axis of maximum tension.

If θ is chosen so that T'_1 is a maximum, we find

$$\tan 2\theta = \frac{2T_6}{T_1 - T_2} \quad (21)$$

and

$$\begin{aligned} T'_1 &= \frac{T_1 + T_2}{2} + \frac{1}{2} \sqrt{(T_1 - T_2)^2 + 4T_6^2} \\ T'_2 &= \frac{T_1 + T_2}{2} - \frac{1}{2} \sqrt{(T_1 - T_2)^2 + 4T_6^2} \end{aligned} \quad (22)$$

⁷ Coker and Filon, Photoelasticity, Cambridge University Press, 1931. M. Hetenyi, Handbook of Experimental Stress Analysis, John Wiley and Sons, Chap. 17, 1950. R. D. Mindlin, J. Applied Physics, April 1939, pp. 222-241 and May 1939, pp. 273-294. W. P. Mason, Electrooptic and Photoelectric Effects in Crystals, Bell System Tech. J., **29**, pp. 161-188, April, 1950.

and hence

$$T'_1 - T'_2 = \sqrt{(T_1 - T_2)^2 + 4T_6^2}. \quad (23)$$

Hence, the birefringence is directly proportional to the difference between the principal stresses.

The retardation R is the difference between the path length for the fast and slow rays when they are transmitted through the thickness of the specimen. If h is the thickness of the specimen

$$h = v_2 t, \quad h - R = v_1 t \quad (24)$$

and hence

$$\frac{R}{h} = \frac{v_2 - v_1}{v_2} \quad \text{and} \quad \frac{c}{v_1} \frac{R}{h} = \left(\frac{v_2 - v_1}{v_2} \right) \frac{c}{v_1} = \left(\frac{c}{v_1} - \frac{c}{v_2} \right) = B. \quad (25)$$

Hence, the retardation is given by

$$\frac{R}{h} = \frac{C'c}{v_1} (T'_2 - T'_1) = C(T'_2 - T'_1). \quad (26)$$

If λ is the wavelength of light used for the measurement,

$$\frac{N\lambda}{h} = \frac{R}{h} = C(T'_2 - T'_1), \quad (27)$$

or the number of fringes is related to the difference of the principal stress by

$$N = \frac{Ch}{\lambda} (T'_2 - T'_1). \quad (28)$$

For photoelastic bakelite for the green mercury line 5461 Å, the fringe constant λ/C is 88 pounds per square inch/inch/fringe. Hence the difference between the principal stresses is given by

$$(T'_2 - T'_1) = \frac{N}{h} (88) \text{ pounds/sq in.} \quad (29)$$

The isochromatic lines such as shown on Fig. 3 are taken with quarter-wave plates in addition to the crossed polaroids. These lines give directly the multiple number of times the stress is greater than the value $(88/h)$ pounds per square inch. In order to know the exact multiple, one has to know the starting point or the points of zero stress difference, called isotropic points. These are determined by using white light and locating the gray fringes. The zero order fringe is gray because for this case, all the wavelengths of light are delayed the same amount and hence no

color effects appear. A first order fringe will have red (absence of violet) nearest the zero order fringe and violet (absence of red) furthest from the zero order fringe. High order fringes will not appear at all in white light since they are the resultant of a number of colors which add up to white light.

Having located the zero order fringe, a simple count will give the number of times the factor $(88/h)$ has to be multiplied by to obtain the number of pounds per square inch. This stress will, however, only be a stress difference and in order to resolve this into stresses along X and Y axes and a shearing stress in the XY plane, other information is necessary.

One part of the information is obtained when the isoclinic directions are obtained. These directions are the directions of the principal stress axes and these are obtained by taking out the quarter wave plates and rotating the axes of the polaroids (keeping them crossed) until the polarization axes coincide with the principal stress axes at any point. When this occurs, the picture will be black because if no model were there, the polarized light passed by the polarizer would be cancelled by the analyzer. If the principal axis of the stress ellipsoid coincides with the direction of polarized light from the polarizer, only one ray will be generated whose plane of polarization coincides with that of the polarized light and hence this will be cancelled in the analyzer. The isoclinic lines show up much better if a white light source is used. Hence, the isoclinic lines locate the direction of the principal axes of the stress ellipsoid. From equations (20) and (23) we have

$$T_6 = \left(\frac{T'_1 - T'_2}{2} \right) \sin 2\theta, \quad T_1 - T_2 = (T'_1 - T'_2) \cos 2\theta. \quad (30)$$

Hence, if we know θ (the direction of the principal stress axes with respect to the axes for which the stresses are to be analyzed) the shearing stress T_6 and the difference between T_1 and T_2 can be obtained from equation (30).

The other necessary relation can be obtained from the equilibrium stress relations that have to be satisfied by any stationary body, namely

$$\frac{\partial T_1}{\partial x} + \frac{\partial T_6}{\partial y} = 0; \quad \frac{\partial T_2}{\partial y} + \frac{\partial T_6}{\partial x} = 0. \quad (31)$$

Integrating these equations

$$T_1 = T_{1_0} - \int \frac{\partial T_6}{\partial y} dx; \quad T_2 = T_{2_0} - \int \frac{\partial T_6}{\partial x} dy. \quad (32)$$

For example, the isoclinic lines for the square plate of Fig. 3 are shown by Fig. 24. All the isoclinic lines converge on the positions of the "eyes" which is a characteristic to be expected of an isotropic point ($T'_1 - T'_2 = 0$). For the line through the center of the "eyes", which we designate as the X direction, the isoclinic line lies along the X axis and hence there is no shearing stress along this axis and $T_1 - T_2 = T'_1 - T'_2$. At either edge the stress in the X direction is zero and hence at this point the stress T_2 (which is a compression) is obtained by counting the number of lines from the isotropic point to the edge (in this case 20) so that the stress at the edge is

$$T_2 = \frac{20 \times 88}{0.95} = 1850 \text{ pounds/square inch} \quad (33)$$

To determine the shearing stresses at any point, one needs to fit the isoclinic lines over the isochromatic lines. For example, for any point along the isoclinic line marked 15, the direction of the principal axis is

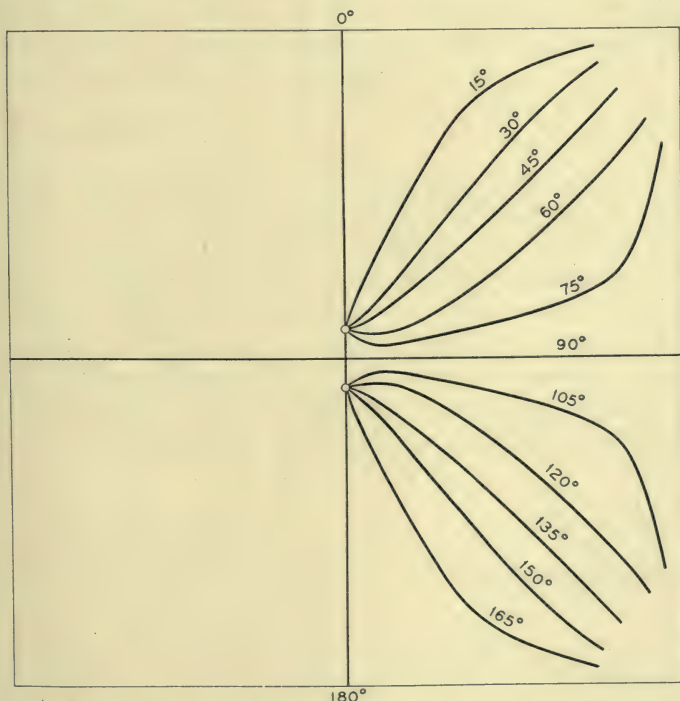


Fig. 24 — Isoclinic lines for a square terminal whose isochromatic lines are shown by Fig. 3.

15° from the X axis, since it was obtained by setting the analyzer and polarizer at 15° from the direction X . Counting the number of isochromatic lines from the "eye", the principal stress difference $T'_1 - T'_2$ is known and using equation (30), T_6 is at once calculated. To obtain $\partial T_6/\partial y$, it is sufficient to divide T_6 by the distance y that the point is above the X axis. Proceeding this way, values of $\partial T_6/\partial y$ are plotted as a function of X . Then from the first of equations (32) with $T_{10} = 0$ at the edge, one can obtain T_1 by integrating $\partial T_6/\partial y \, dx$ over the X axis. The result is the curve marked T_1 of Fig. 5. Having T_1 , T_2 can be obtained from the isochromatic lines which determine $T_2 - T_1$ and this is plotted on the upper part of Fig. 5.

Solderless Wrapped Connections

PART III — EVALUATION AND PERFORMANCE TESTS

By R. H. VAN HORN

(Manuscript received February 9, 1953)

In the development of solderless wrapped connections the basic requirements of electrical and mechanical stability have been translated into test requirements on laboratory samples of these connections and on the manufactured product. These tests have shown that the connections can withstand satisfactorily the effects of corrosion, humidity, vibration, and relaxation. The effects of terminal dimensions, materials, corner sharpness, wrapping tool construction, etc. are noted.

INTRODUCTION

The previous two articles have described the fundamental considerations involved in solderless wrapped connections. A description of these connections together with a rather detailed explanation of the forces which maintain them has been presented. This third article discusses the results of a number of tests where the actual fabrication and use of these connections have been simulated. From these results it will be seen that reliable performance can be expected over the central office life of these connections, and that the variations permissible in their fabrication will provide sufficient margin to make that process easy to control.

GENERAL REQUIREMENTS

The minimum physical requirements which a solderless wrapped connection must meet are:

1. Intimate contact between wire and terminal.
2. The points of contact should be gas-tight to withstand corrosion.
3. The minimum dimension of the gas-tight area should be great enough so that it does not decrease appreciably during the expected

life (forty years) because of corrosion or because of relaxation of the internal stresses in the wire or terminal.

4. The sum of the areas of intimate contact should be equal to or larger than the cross-section of the wire to prevent local heating.

5. The connection should be mechanically stable so that forces applied to the connection during shipment, installation and subsequent maintenance activities will not dislodge the wire and break the points of intimate contact.

6. The wire should not be embrittled during the wrapping operation so that it will subsequently break due to vibration, handling, or unwrapping.

TRANSLATION OF GENERAL REQUIREMENTS INTO TEST REQUIREMENTS

In order to evaluate the feasibility of solderless wrapped connections, extensive development studies were necessary so that a good estimate could be obtained as to whether these connections would meet these general requirements under a wide variety of conditions and with sufficient margin to provide for ease of manufacture. It was necessary to translate these requirements first into a set of development test requirements and second into a set of shop inspection requirements. These two translations of requirements will not necessarily be the same although there will be a large degree of overlap. In either translation they can be broken down into two distinct areas. These two areas of tests cover those tests which are related to evaluating (1) life and electrical stability of the connection and (2) mechanical stability.

A great deal of engineering judgment was used in the translation of the physical requirements into inspection requirements and this judgment took into account the special nature of conditions to be encountered in telephone offices. Consideration was given to the methods of handling the wire, the manipulations of installation and maintenance men when working on central office equipment, the effect of vibration, the effect of variation in tool dimensions and the like. Furthermore, a good deal of knowledge of the corrosion and relaxation process had to be developed before it could be judged on the basis of accelerated tests that the connections might be expected to have a satisfactory field life.

There may be applications where the translation of the physical requirements into test requirements may be different, perhaps quite different, from the translation made for the telephone apparatus which was in mind during this investigation. The use of other kinds and sizes of wire or terminals may require an evaluation quite different from the one presented here. Nevertheless, the test requirements herein estab-

lished should have a wide application in many areas. In telephone practice they provide a reasonable latitude for variations in the process of making the connection, including tolerances in the parts, and at the same time guarantee a good product.

Most of the product tests which have been made so far apply specifically to connections which use Standard No. 24 tinned solid copper wire such as is used in 95 per cent of the telephone switching plant, and terminals having a rectangular cross-section punched from sheet nickel silver, brass or bronze and whose dimensions are one-sixteenth inch wide by the thickness of the stock (0.013" to 0.062"). These terminals are typical of those which are or could be used on relays, switches, resistors, capacitors, terminal strips, etc. Studies with #20 and #22 wire have been made with results similar to those with #24 wire.

A similar investigation is under way on connections to terminals which are made from round silicon copper and nickel silver wire such as are used on the wire spring relay, and where the wire terminal had been prepared for connection by various treatments such as flattening, coining, serrating, annealing, etc.

AGING OF WRAPPED CONNECTIONS

Assuming that a connection can be wrapped with sufficient mechanical strength to withstand handling, vibration, etc., there appear to be two factors which might cause the connection to fail after a period of time. These factors are (1) relaxation of the internal stresses in the metals, and (2) corrosion of the metal surfaces. As Mr. Mason points out in his paper, it now appears that solid state diffusion of metal across the boundary between the wire and the terminal improves the connection as much or more than relaxation degrades it. Tests have been designed to relax the connections in a short time to the same degree that will occur in the normal forty-year life which is required. These tests will be described later.

Several investigators have studied the rate of surface corrosion in indoor atmospheres of metals such as are used in these connections.

Studies of corrosion* where oxidation is the primary factor indicate for example that the corrosion in zinc varies linearly with time and on copper it varies as the square root of the time exposed. The corrosion products are ZnO and Cu_2O . The corrosion rate for brasses fall between

* British Non-ferrous Metals Research Association, Investigation on the Atmospheric Corrosion of Non-ferrous Metals, First and Second Experimental Reports to the Atmospheric Corrosion Resistance Committee, May, 1924, and May, 1927, W. H. J. Vernon.

TABLE I—THICKNESS OF MATERIAL WHICH CORRODES IN FORTY YEARS

Zinc	0.00023"
Copper	0.000105"
Tin	Negligible

copper and zinc. There are very few data available on tin but its corrosion should be less than either copper or zinc. From the data presented the depth of metal which will corrode during a forty-year period in a central office for several metals is estimated* to be as shown in Table I.

Samples of copper bus-bar taken from telephone exchanges where they have been exposed for periods up to forty years show that the tarnish is primarily Cu_2O and the actual rate of corrosion may be appreciably less than that estimated by the above figures.

Thus the depth of metal corroded is small enough to neglect when the metals are subject to indoor atmospheric exposure. When dissimilar metals are joined in a connection there is the possibility of electrolytic corrosion in addition to atmospheric corrosion. The particular metals involved here, however, are relatively close to each other in the electromotive force series of metals so that it is expected that this effect will be negligible especially as there is no condensation on these connections.

It is therefore expected that the most important factor in the aging of these connections is the relaxation of stresses internal to the wire and the terminal rather than corrosion. The test procedures and results in the following sections reflect that view.

DEVELOPMENT TEST PROCEDURE — ELECTRICAL REQUIREMENTS

General Requirements 1 through 4 are considered together. A set of tests has been designed to evaluate the degree to which these requirements can be met.

Since the tests which have been devised are destructive it is necessary to check connections in production on a sampling basis.

In determining whether a connection meets the General Requirements 1 through 4, Test Procedure I as follows was devised.

Test Procedure I for Solderless Wrapped Connections

1. Check connection for insulating barrier film between wire and terminal.

2. Measure the variation of the resistance of connection while producing movement between the terminal and connecting wire.

* Unpublished memoranda, D. H. Gleason, Bell Telephone Laboratories.

3. Chill for two hours at 0°F.

4. Heat for three hours at a temperature (175°C) which will relax the stresses as much as will occur during the expected life at the normal central office operating temperature.

5. Expose the connection to a suitable agent which will discolor all the non gas-tight area.

6. Remeasure the resistance variation as in 2.

7. Unwrap the wire and estimate the size of the gas-tight areas.

Items 1 and 2 are intended to show whether initially there is the intimate contact between wire and terminal demanded by the General Requirement No. 1. At present, the multiple terminal banks on step-by-step switches are connected together with a clinched solderless connection. Based on experience with these connections, together with an estimate of the noise produced (See appendix A) a resistance variation in excess of 0.002 ohms is considered to indicate a poor connection. Items 5 and 7 will show the existence of the gas-tight areas demanded by General Requirement No. 2, and the size estimate from Item 7, the area of contact demanded by General Requirement No. 4 can be evaluated. Items 3 and 4 are intended to simulate most of the aging which will take place during the life of the connection. It is believed, as described earlier in this paper, that relaxation of internal stress is the chief factor in the aging of these connections. The chill at 0°F puts the maximum initial stress in the wire by shrinking it at extreme operating

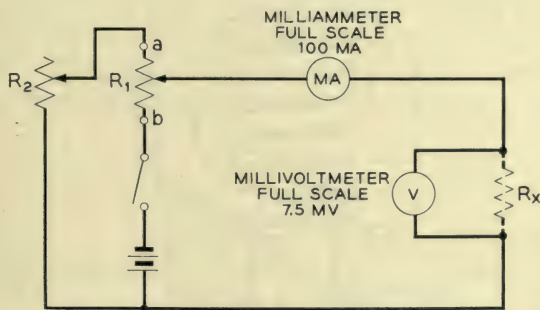


Fig. 1 — Test set for measuring resistance of solderless connection.

TEST PROCEDURE

- (1) With R_1 set at "a" and switch SW closed, adjust R_2 so that the millivoltmeter reads 25 microvolts with R_x open.
- (2) If millivoltmeter returns to zero with test connection across R_x , no barrier film is indicated and therefore connection is closed.
- (3) All resistance measurements are made with R_1 set to give 100 milliamperes through R_x . Then $R_x = \text{millivolts}/0.100$.
- (4) A variation of not more than 2 milliohms as the connection is moved or disturbed indicates a stable connection.

temperatures. The heating for three hours at 175°C produces the degree of relaxation expected over a 40-year period. By examining and estimating the size of the gas-tight areas after this process, the necessity for maintaining the intimate minimum gas-tight area demanded by General Requirement 3 is considered to be met. Item 6, the remeasurement of resistance is further confirmation of proper performance after relaxation.

For the measurement of resistance, and barrier films a circuit is used as shown on Fig. 1. With R_1 set at "a," R_2 is adjusted so that the voltage of about 25 microvolts is applied across R_x before the connection to be measured is inserted. This voltage is low enough to insure the absence of a film and at the same time gives a convenient reading on the test set. If the millivoltmeter drops to approximately 0 when the connection is inserted it indicates that no barrier film exists at the connection. The current is then increased to 100 milliamperes by means of the potentiometer, R_1 , and the resistance is determined from the millivoltmeter reading. Since most of the measured resistance is in the wire rather than in the connection the important criterion of quality is the variation in resistance as the wire is moved relative to the terminal. If the variation in resistance of the connection does not exceed two milliohms when the wire is moved back and forth the connection is considered to be

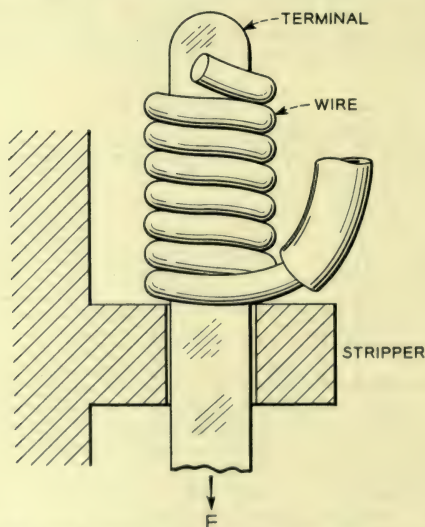


FIG. 2 — Stripping force test for solderless connection. (1) The connection shall consist of six turns minimum of which at least four are close wound. (2) The connection shall be capable of withstanding a stripping force, F , of at least 3000 grams applied as shown. (3) The wire shall be capable of being unwrapped from a terminal without breaking.

stable and the requirement for intimate contact between wire and terminal is considered met. In a typical local talking circuit this amount of resistance variation would correspond to a noise level of approximately -8 db where 0 db equals 10^{-12} watt and where anything less than $+26$ db would give no noise transmission impairment (See Appendix A).

DEVELOPMENT TEST PROCEDURE — MECHANICAL REQUIREMENTS

In determining whether a connection meets the General Requirements 5 and 6, Test Procedure II as follows was devised:

1. The connection shall be capable of withstanding a stripping force of 3000 grams applied as shown in Fig. 2.
2. The wire shall be capable of being unwrapped from a terminal without breaking.

These items are related directly to General Requirements 5 and 6. Experience has shown that under ordinary conditions of handling of connections as cables and frames are wired in the shop, or when a second connection is being made on an already wired terminal, a resistance to stripping in excess of about 3000 grams is required if the demands of General Requirement 5 is to be met. To insure adequate mechanical strength and current carrying capacity, a minimum of six turns is required. If the turns are close wound the strip-off force is readily met. If they are open wound they may strip off with a much lower force.

When the edge radius R (Fig. 3) of the wrapping tool is too small very high tension can be developed in the wire while wrapping. At very small radii the tension can be sufficient to break the wire. Although wrapping with high tensions in the wire produces a connection which will sustain very high stripping force, the wire is so embrittled that

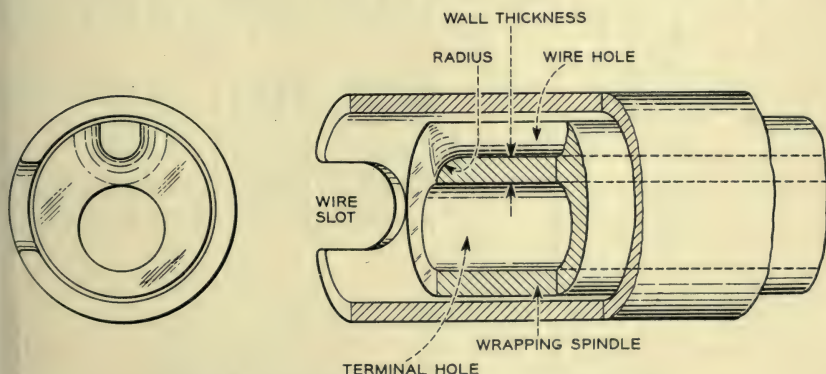


Fig. 3 — Solderless connection wrapping tool.

TABLE II—SOLDERLESS WRAPPED CONNECTION HUMIDITY AND CORROSION EXPOSURE TESTS

Test Specimens					Exposure					Result	
Terminal	Wire	No. of Wraps	Average Stripping Force	No. of Samples	Vibra- tion	0° F.	180° F.	Corrosive Agent	Hours In Corro- sive Agent	Months at 85°F 90% Rel. Humidity	Number of Connections Where Resistance Exceeded 0.002 Ohms
Untinned brass	<i>Inches</i> 0.030 × 0.090 0.040 × 0.040 0.020 × 0.060 0.030 × 0.030 0.040 × 0.090	Size 24 tinned 24 tinned 24 tinned 24 tinned 24 tinned	6 6 6 6 6	High† High High High High	24 60 60 60 48	<i>Hours</i> 50* 50 50 50 50	<i>Hours</i> 2 2 2 2 2	<i>Hours</i> 2 2 2 2 2	<i>Hours</i> 1½ 1½ 1½ 1½ 1½	<i>Month</i> 21 7½ 7½ 7½ 21	None None None None None
Brass-nickel plated and tinned	0.013 × 0.062 0.013 × 0.062 0.013 × 0.062 0.013 × 0.062	24 tinned 24 tinned 24 tinned 24 tinned	3 4 5 6	Low† Low Low Low	25 25 25 25	0 0 0 0	2 2 2 2	2 2 2 2	<i>Hours</i> 1½ 1½ 1½ 1½	9 10½ 10½ 10½	25 5 1 None
Brass-nickel plated and tinned	0.013-0.031 × 0.047 0.031 × 0.062 0.031 × 0.047 0.031 × 0.062 0.013 × 0.125 0.031 × 0.125 0.013 × 0.125 0.031 × 0.125 0.031 × 0.125	24 tinned 24 tinned 24 tinned 24 tinned 24 tinned 24 tinned 24 tinned 24 tinned 24 tinned 24 tinned	6 6 6 6† 6 6 6 6 6 6	High† Low† Low 5000 grams 3100 grams 3400 grams 2100 grams 2600 grams 1800 grams	60 25 20 25 10 10 10 10 10 10	0 0 0 0 50 50 50 50 50 50	2 0 0 0 2 2 2 2 2 2	2 0 0 0 2 2 2 2 2 2	<i>Hours</i> 1½ 0 0 0 1½ 1½ 1½ 1½ 1½ 1½	15 9 9 9 2 2 2 2 2 2	None None None None None None None None None None

Untinned nickel silver	0.035 × 0.125	22 tinned	4	4000 grams	5	0	2	2	HCl	$\frac{2}{3}$	10	None
	0.035 × 0.125	22 tinned	4	4000 grams	5	0	2	2	SO ₂	$\frac{2}{3}$	10	None
	0.035 × 0.125	22 tinned	4	4000 grams	10	0	2	2	—	0	10	None
	0.040 × 0.125	22 tinned enameled	4	5400 grams	10	0	2	2	H ₂ S	$\frac{1}{2}$	6½	None
	0.040 × 0.125	20 tinned enameled	4	Hight†	10	0	2	2	H ₂ S	$\frac{1}{2}$	6½	None
Steel-tin plated	0.035 × 0.125	22 tinned	4	Low†	50	0	2	2	H ₂ S	$\frac{1}{2}$	6½	None
Wire spring relay single wire	Flattened and solder dipped	24 tinned	6	3700 grams	25	0	2	2	H ₂ S	$\frac{1}{2}$	12	None
	Diamond shaped and solder dipped	24 tinned	6	4400 grams	25	0	2	2	H ₂ S	$\frac{1}{2}$	12	None
Twin wire	Solder dipped only	24 tinned	6	3700 grams	25	0	2	2	H ₂ S	$\frac{1}{2}$	12	None
	Diamond shaped and solder dipped	24 tinned	6	4400 grams	25	0	2	2	H ₂ S	$\frac{1}{2}$	12	None

* This was the time at which similar soldered connections began to show breakage.

† Estimated from wrapping tool dimensions.

‡ One turn insulated, five turns bare.

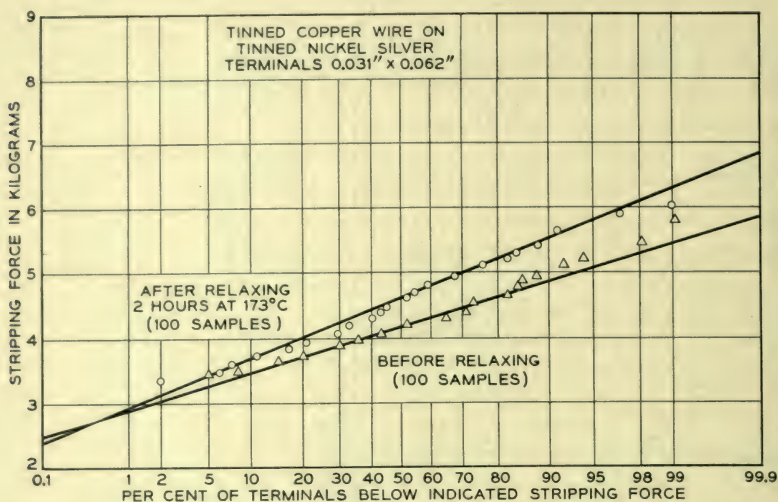


Fig. 4 — Effect of relaxation on stripping force.

under vibration and handling it may break easily. A practical and generally easily met requirement is that the connection be capable of withstanding unwrapping without wire breakage.

OTHER TESTS

A large amount of work has been done other than testing connections in accordance with the foregoing requirements. These tests (See Table II) include measurement of exposure to humidity, hydrogen sulphide corrosion, heating by internal electrical losses, vibration, etc. Many of these tests were made before any clear picture mechanism of deterioration of the connection was available so that they were not exposed to temperature high enough to cause very much relaxation. Therefore, there is no quantitative relation between these tests and the actual aging of connections, but they do indicate a general ruggedness which gives confidence in the reliability of the connections.

In measuring the effect of relaxation, connections have been tested in a preliminary way by subjecting them to sufficient heat sufficiently long to relax the stresses appreciably more than would be expected in a normal forty-year life. Upon unwrapping, the bright areas of contact were still present. Fig. 4 shows the result of a typical relaxation test on the stripping force. The rise in stripping force after relaxation is attributed to the solid state diffusion process described by Mr. Mason.

In order to compare how the solderless wrapped connection and soldered connections are able to withstand the effects of vibration over a period of time, several sets of terminals were mounted on a test table and wires from a cable were wrapped or soldered to the terminals. The cable was then oscillated through a small angle at 19 cycles per second and the elapsed time until lead breakage occurred was recorded. Some early similar tests indicated that soldered connections began to break at about fifty hours and the solderless connections lasted several times that long. In a more recent test where ten connections of each type were set up three of the soldered connections broke within six hours and all soldered connections broke within 1,000 hours. None of the wrapped connections had broken in that interval. In another test using ten soldered connections and twenty solderless connections, one of each broke at nineteen hours and after fifty-seven hours two more soldered connections broke with no further breakage of solderless connections. Further tests are being made of resistance to vibration under various conditions but the early indications are that the life of the solderless connections is as good as and perhaps better than that of the soldered connections.

The increased resistance of the wrapped connection to vibration breakage comes about because in the soldered connection the bending stresses are concentrated at the point of emergency of the wire from the solder, while in the wrapped connection the stresses tend to be distributed over the entire first turn of the connection. In order to obtain long life, it is important to avoid any exposed nicks in the wire at the first turn of the connection. Otherwise stress concentrations, perhaps as severe as those in soldered connections may be set up.

Incidentally, unless the wire has an enamel or similar coating it is unnecessary to clean the wire or the terminal before wrapping, since forces set up as the wire wraps on to the terminal shear off surface films and the resultant gas-tight areas are immediately established on clean surfaces. It is this intimate contact which also makes possible the diffusion process between the wire and the terminal.

A large number of parameters have an effect on the level of stripping force which a given connection will withstand. Among these are the number of wraps, tension on the wire during wrapping, sharpness of the corners of the terminals, the elastic properties of both the terminal and the wire, the size of the terminal, the hardness of the terminal and wire, the presence or absence of tin plating, taper of the terminal, aging, tool design, etc. Tests have been run showing the effect of variations in these parameters on the stability of the connections. Typical results of these

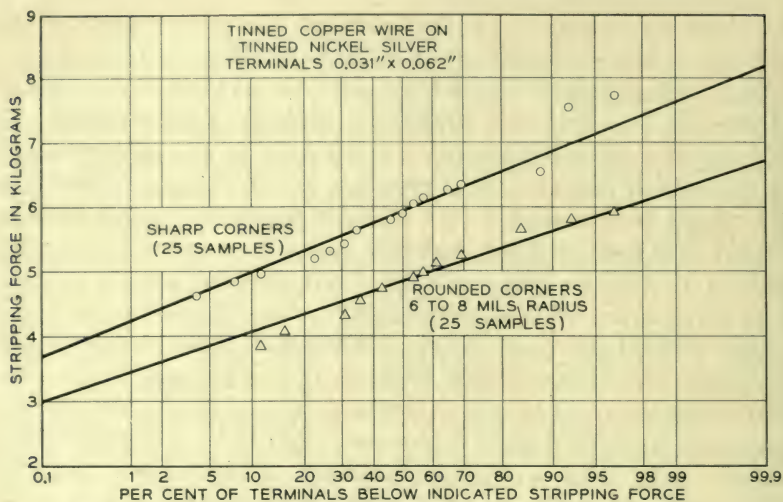


Fig. 5 — Effect of terminal edge sharpness on stripping force.

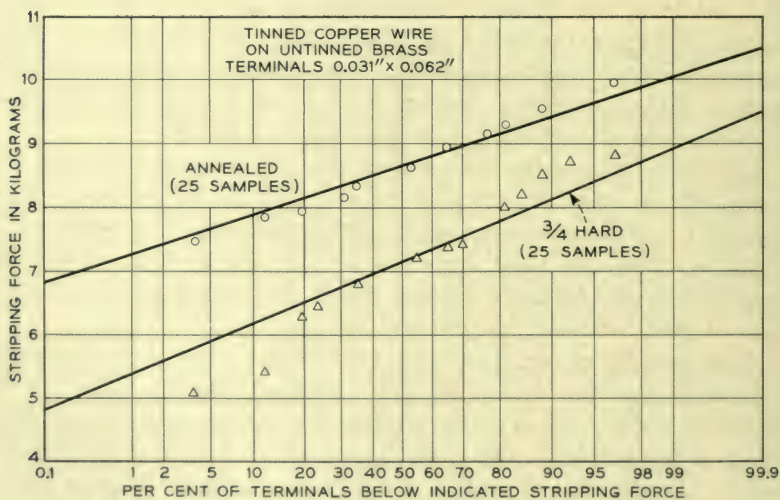


Fig. 6 — Effect of terminal hardness on stripping force.

tests are shown on Figs. 5 to 9 inclusive. These charts show a plot on arithmetic probability paper of stripping force against the cumulative per cent of the sample below the indicated stripping force. The straight lines shown there are actually drawn through the average and average minus 3σ points where σ is the standard deviation of the observed readings. In the cases shown, and in fact in practically all the tests made to date the results have come out so that the experimental values fit very well on the straight line drawn through these points, thereby indicating good normal statistical distributions of the data.

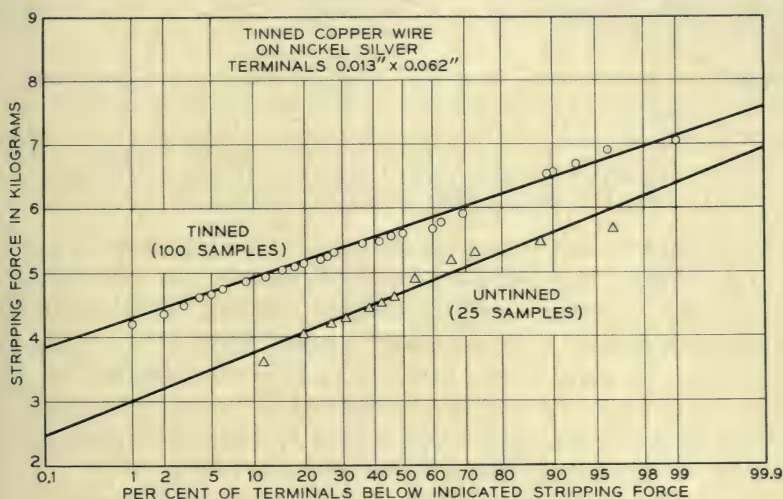


Fig. 7 — Effect of terminal plating (tin) on stripping force.

For example, Fig. 5 shows the effect of stripping force of rounding the corners of the terminal. These curves are for a typical tinned nickel silver relay terminal 0.031" by 0.062" in crosssection. The terminals represented by the upper curve were punched from sheet nickel silver stock and the corners were sharp as is usually the case where shearing operations are used. The lower curve represents connections where the corners were rounded off to a 0.006" to 0.008" radius. It is seen that on the average the stripping force using the sharp corners is about 1,000 grams higher than with the rounded corners, although both sets of terminals met the 3,000 gram requirement for mechanical stability.

The effect of terminal hardness on stripping force is shown in Fig. 6. Here brass terminals were used and it is evident that the softer terminals give considerably higher stripping force than the harder ones. Both

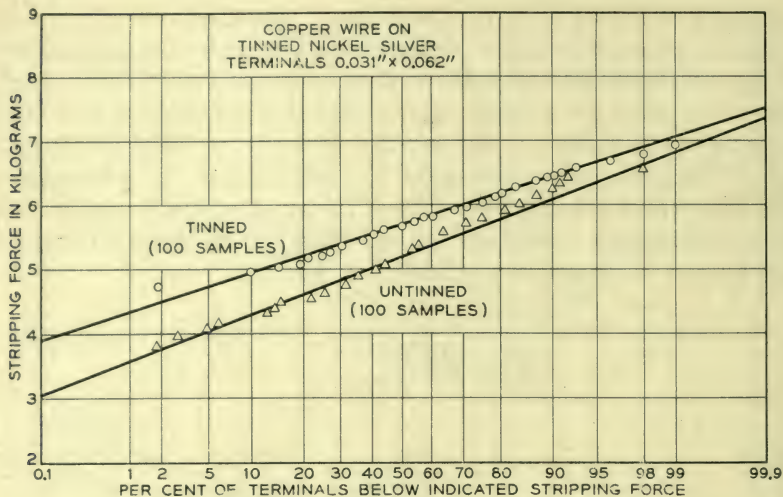


Fig. 8 — Effect of wire plating on stripping force.

curves run higher than the curves for the harder nickel silver in Fig. 5. Fig. 6 shows that the presence of tin plate on a 0.013" by 0.062" terminal increased the minimum expected stripping force by more than 1,000 grams. Figure 8 shows that on a 0.031" y 0.062" terminal that while the effect of the tin plate on the wire was to raise the stripping force as before, the increase was not as great as that in Fig. 7. It is also clear that the thicker terminal of Fig. 8 results in a higher stripping force than that for the corresponding thinner terminal of Fig. 7.

In analyzing the effect of tool variations it has been convenient to take a series of curves like those in Figs. 5 to 8, and from each curve select two points, namely one for the average stripping force and one for the average minus 3σ where σ is the standard deviation of the observed values of stripping force. Using these two points for each of several sets of curves a plot may be obtained of the average and minimum expected stripping force as one or another of the parameters of the tool design are changed. In this manner curves of the type shown on Fig. 9 were obtained. By preparing curves of this type for any given application, the permissible range, Z , of variation of tool radius, R , can be obtained. This range includes those values of R which are above that value which results in wire breakage and below that for which the minimum expected stripping force falls below the stripping force limit.

Table II summarizes the results of many of the tests which have been run on a large number of samples in which the materials and dimensions of the terminals, wires, tightness of wrap, etc., were varied. All of sample

connections originally met the initial test requirements outlined above for resistance variation. They were then subjected to the indicated exposures including being kept in the Humidity Room at 85°F and 95 per cent relative humidity for periods up to two and one-half years, following which they were remeasured. In no case was any barrier film found which did not break down at 25 microvolts nor was any resistance variation in excess of 0.002 ohms observed. Some of these terminals are shown in the photographs of Figs. 10 and 11. Upon unwrapping some of these connections there were still considerable areas which were bright (See Fig. 12), indicating that they were still gas-tight. This would seem to indicate a life expectancy of many years. Since connections

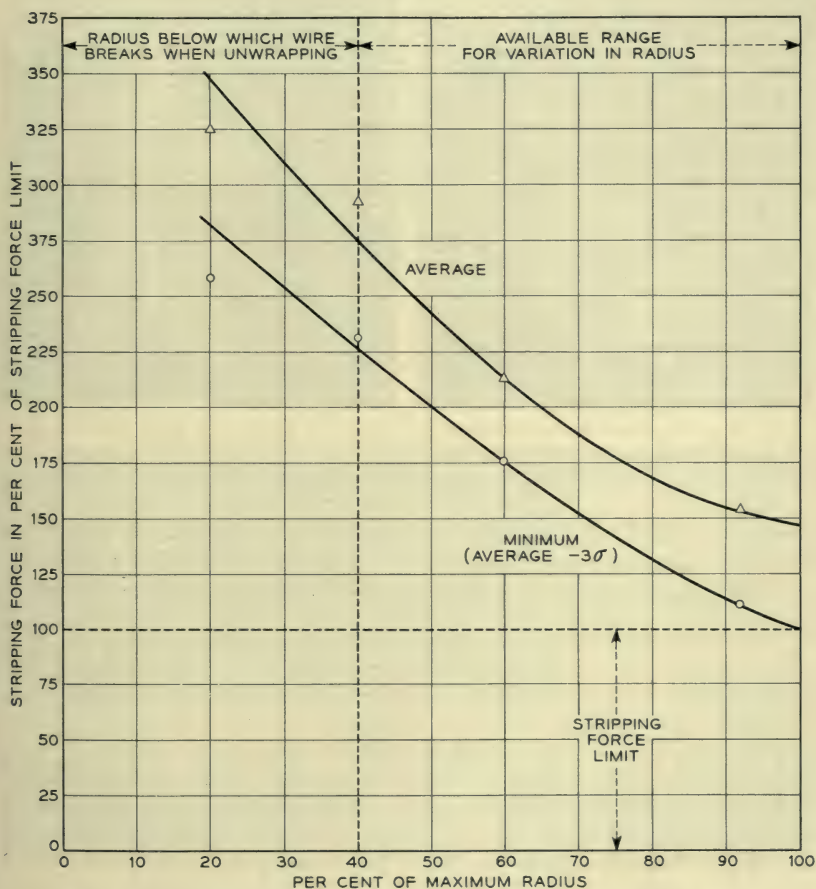


Fig. 9 — Stripping force as a function of radius of wrapping tool.

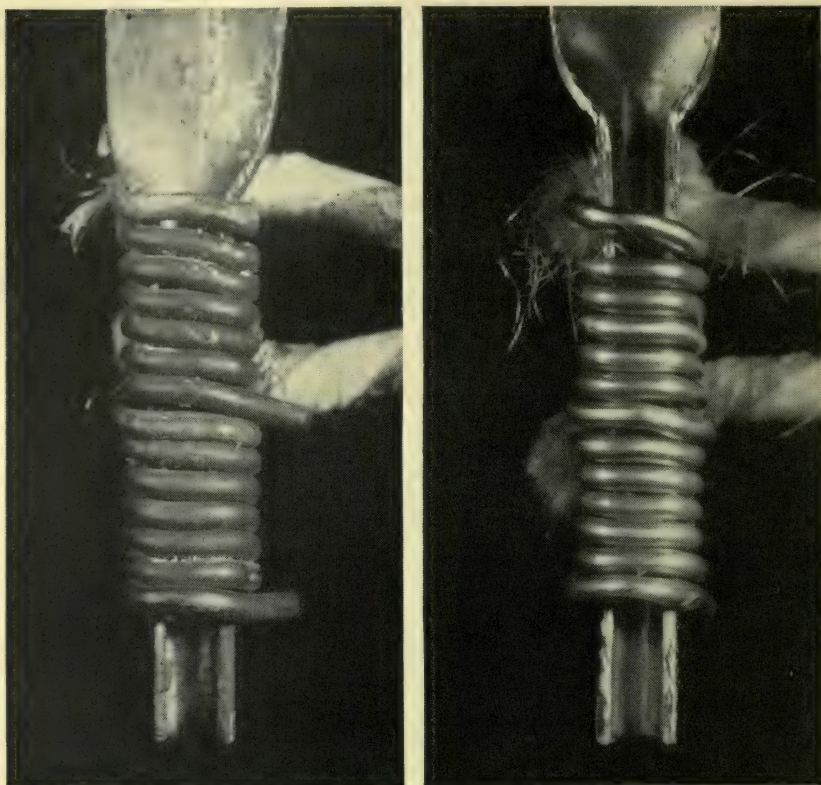


Fig. 10 — Terminal at left exposed to H_2S for 14 hours and then 30 months at $85^\circ F$ 90%RH. The connections did not develop resistance variation in excess of the 0.002 ohm limit. Corresponding new terminal is shown at right.

which meet the 3,000-gram strip-off requirement appear satisfactory on the other tests, the strip-off test used on a sampling basis has been standardized as the shop control of the wrapping tool and indirectly of the quality of the connections themselves.

FIELD TRIALS

A number of equipment units using these connections is currently on field trial. At Tonawanda, N. Y., one trunk unit using wire spring relays of an early design and 300 solderless wrapped connections has been in service successfully since 1951. At Elmhurst, L. I., a sender frame comprising five senders with about 7,500 connections has been in use about a year and a half with no troubles reported to date. At Boston, Mass., an outgoing sender frame for the No. 4 crossbar office comprising three

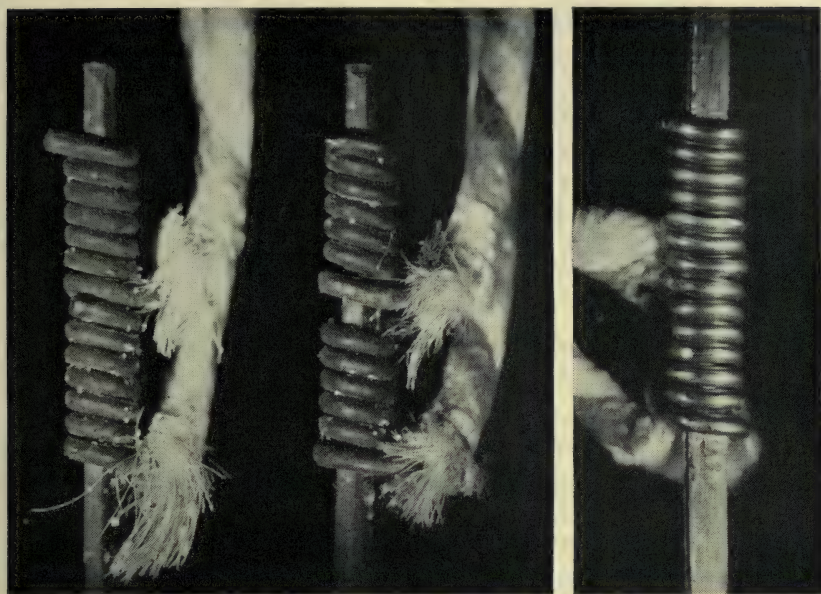


Fig. 11 — Terminals on left exposed to H_2S for one-half hour and $85^\circ F$ 90 per cent RH for seven and one-half months without development of excessive resistance variation. Corresponding new connection is shown on right.

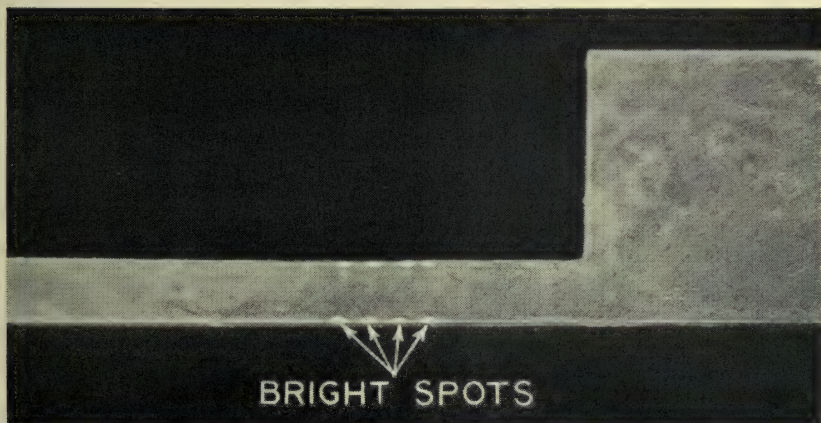


Fig. 12 — Corroded nickel silver terminal after unwrapping connection. Note the bright spots where the wire was in contact with the terminal.

senders with 6,000 solderless connections has been in service about nine months, again with no troubles reported.

In order to test these connections in a transmission circuit one channel of the K-1 carrier system between Newark and Atlanta was selected. Rather than wire a regular equipment unit with solderless wrapped connections, four units were built consisting of two groups of 320 connections in series. Each unit was exposed to 15 weeks at 85°F 90 per cent relative humidity and then inserted in the system. One unit was located at each of the following places: New York, Philadelphia, Baltimore, and Richmond. No troubles or adverse service reactions have been received to date, more than six months after installation.

SUMMARY

The tests described indicate that solderless wrapped connections are practical when wrapping No. 24 solid tinned copper wire on flat punched terminals of brass or nickel silver where the width is one-sixteenth inch and the thickness varies from 0.010" up to one-sixteenth inch. Similar tests with heavier wire, (No. 20 to No. 23) such as would be used in distributing frames and tests with No. 24 wire on flattened, coined, or otherwise treated wire spring relay terminals have also been successful. These connections are mechanically stable, and have less tendency to break due to handling and vibration than solder connections, and will

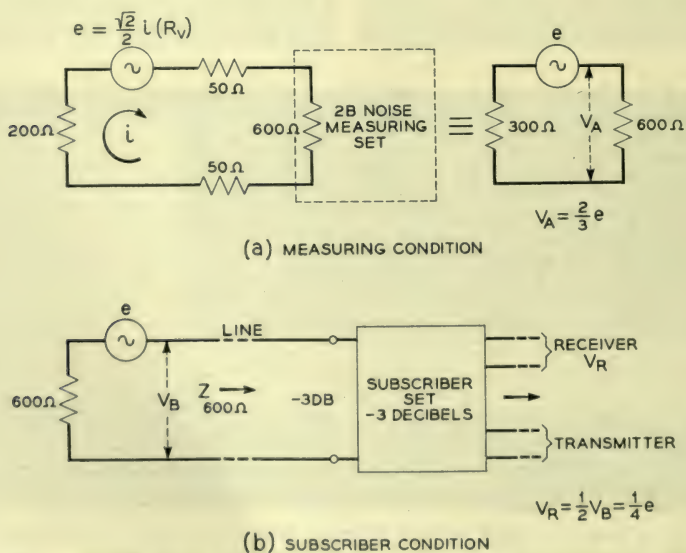


Fig. 13 — Comparison of noise measuring circuit with subscriber circuit.

probably have a central office life of more than fifty years without developing objectionable resistance variation. Tests on other terminals and using wires heavier than No. 20 and on wires smaller than No. 24 are to be made.

ACKNOWLEDGEMENT

The author acknowledges that this paper reflects largely work done under the supervision of D. H. Gleason whose recent retirement prevented his authorship.

APPENDIX A

NOISE LEVEL ARISING FROM VARIABLE CONTACT RESISTANCE

The measurement of contact resistance noise in a telephone central office has been standardized, and the estimate of the effect of a given resistance variation has been made in accordance with the standard technique.

In this measurement, a 2-B noise measuring set is connected as shown in Fig. 13 (a) where the voltage, e , is that due to a variation in contact or connection resistance. If a dc current i is flowing and the resistance variation is $(\Delta R)_0 \sin \omega t$, then

$$e_{\text{eff}} = \frac{\sqrt{2}}{2} i \cdot (\Delta R) \quad (1)$$

and

$$V_A = \frac{2}{3} e. \quad (2)$$

In the subscriber condition (Fig. 13 (b)) there is a 3 db loss in the subscriber's loop and a 3 db loss at the subset so that the voltage at the receiver

$$V_R = \frac{1}{2} V_B$$

and

$$V_b = \frac{1}{2} e$$

It therefore follows that

$$\frac{V_A}{V_R} = \frac{2e}{3} \div \frac{1}{4} e = \frac{8}{3}$$

and

$$20 \log \frac{V_A}{V_R} = 9 \text{ db (to the nearest decibel).}$$

Extensive tests have shown that a level of 17 db at the receiver where 0 db = 10^{-12} watts will produce no noise impairment of transmission. The corresponding level at the input to the noise meter in the measuring condition is $17 + 9$ or $+26$ db.

Currents in talking circuits may vary from 0.025 amperes in many central office circuits to approximately 0.150 amperes in short subscriber's loops. At a current of 0.025 and assuming $\Delta R = 0.002$ ohms the noise voltage will be

$$e = \frac{\sqrt{2}}{2} \times 0.025 \times 0.002 = 35.4 \times 10^{-6} \text{ volts}$$

and the voltage at the input to the noise measuring set will be $\frac{2}{3} e = 23.5 \times 10^{-6}$ volts. The power into the set will be

$$\frac{23.5 \times 10^{-6}}{600}$$

or 1×10^{-12} watts or 0 db. The noise measuring set is equipped with a weighting network which takes into account the relative interfering effect of different noise frequencies on received speech. For random noise, the effect of this network is to reduce the measured noise level by about 8 db as compared with the reading that would be observed without the network. Accordingly the noise produced by the above 23.5×10^{-6} volts would measure -8 db.

An Improved Circuit for the Telephone Set

By A. F. BENNETT

(Manuscript received August 16, 1952)

A telephone set known as the 500 type has been in production for Bell System use for some time.¹ The successful outcome of an intensive study has made it possible to simplify the circuit and some of the components of this set, and thereby to increase dependability and life and significantly to reduce the manufacturing cost. This change now has been completed and telephone sets embodying the necessary modifications are in production. This paper discusses some of the problems involved in this work and outlines the improvements which have been effected. Presented also is information concerning the superior performance of the 500 set over the preceding telephone set when used in noisy locations.

INTRODUCTION

One of the outstanding characteristics of the 500 type set is a 10 db increase in combined transmitting and receiving performance on long loops. This gain is equally divided between receiving and transmitting. This improvement has resulted largely from the use of a transmitter and a receiver which are not only more efficient, but also have better frequency response characteristics. To take full advantage of these transmission gains, two new elements were introduced in the original 500 set design. One of these elements was a better sidetone balancing circuit to offset the more sensitive transmitter and receiver, and maintain sidetone at a level no greater than it was with the previous design of set (known as the 302 type). The other was a tungsten filament and thermistor element to control automatically the transmitting and receiving levels so that the desired gains occur on a graduated basis as the loop length increases. This combination of filament and thermistor bead was called the transmission equalizer. While the transmission objectives were met with telephone sets having these elements, this additional component increased the manufacturing cost of the set appreciably, and therefore more economical means of attaining the desired results were sought. Such means have been found in the form of an arrangement in which a

pair of silicon carbide varistors serves for both the sidetone balancing and equalizing functions, and with certain novel modifications in the circuit of the set, yields essentially the same overall transmission performance as the original design at lower cost. The advantage is retained of a telephone set which from a transmission standpoint is universal in its application in the telephone plant. This universality avoids additional codes of set which require effort and expense to administer. Telephone sets having these features are designated as 500C (manual) and 500D (dial) types. The original design is known as the 500A (manual) and 500B (dial types).

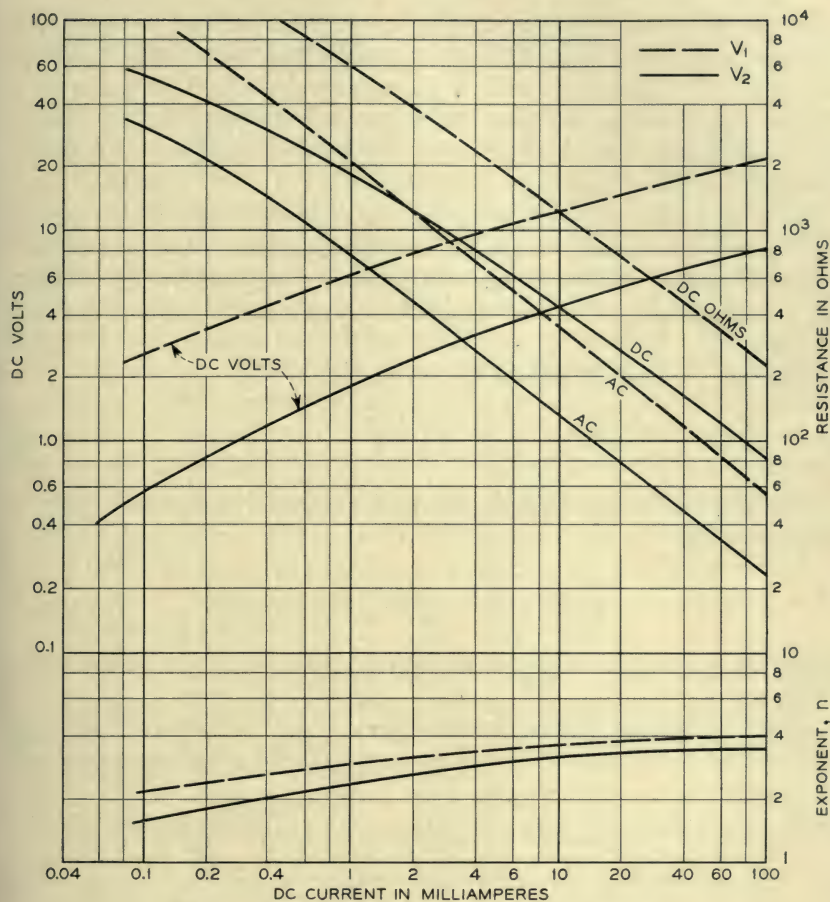
GENERAL

When the sidetone, which is the sound level in the receiver caused by voltage developed in the local transmitter, becomes too high, the subscriber subconsciously lowers his talking level thereby reducing the level of outgoing transmission.^{2, 4} To avoid this loss and make the higher efficiency provided by the new transmitter and receiver effective, a reduction in sidetone must be made. When the room noise level at the station is high, another undesirable effect of high sidetone occurs. This is the loss in effective receiving which results from the masking of incoming speech by the high level of sidetone noise in the receiver. This will be discussed in more detail later.

In developing the original design of 500-type set, several different equalizing means were considered, and the type involving a filament in series with the transmitter, and a thermistor bead in shunt with the receiver was selected as the most suitable with the means available at that time. Reduction in sidetone was realized by the use of a special balancing network which required an auto transformer. The set now being produced is based on a new circuit arrangement, utilizing a pair of semiconductors in the form of silicon carbide varistors, one of which has a resistance in a range which required development for this specific purpose.

CHARACTERISTICS OF SILICON CARBIDE VARISTORS

Fig. 1 shows the relation between the dc voltage and current for the two varistors used in the 500C and D sets. These varistors were coded 312D and 312E, but for brevity in this paper, they are referred to as V_1 and V_2 respectively. Also shown in Fig. 1 is the dc resistance of these varistors as determined by the ratio of applied voltage to current. It is

Fig. 1 — Characteristics of V_1 and V_2 varistors.

observed that the resistance decreases markedly with increasing voltage or current.

The direct current characteristics of a silicon carbide varistor may be expressed in a simple form by

$$I = AE^n, \quad (1)$$

where I is the direct current through the varistor and E is the applied dc voltage. When $n = 1$ the equation (1) reduces to an ohm's law relationship. A and n are generally regarded as constants when considered over a limited range in current. However, over the range shown in Fig. 1, it may be seen from the curvature of the voltage — current char-

acteristic that this is only a rough approximation and can be considered as valid only over a limited range in current. The value of the exponent n , which on a logarithmic scale is the reciprocal of the slope of the voltage-current characteristic and which also is plotted on Fig. 1, is seen to increase appreciably with increasing current. For example, n for the V_1 varistor covers a range from about 2 to 4.

In a telephone set the alternating signal currents are small relative to the direct currents. Therefore, in explaining how the silicon carbide varistor controls the transmission equalization and sidetone, it is necessary to examine how its ac resistance to small signals varies with direct current.

Equation (1) can be put in the form

$$E = \left(\frac{I}{A} \right)^{1/n} \quad (2)$$

Differentiating this with respect to I we get for the ac resistance to small signals

$$\frac{dE}{dI} = \left(\frac{1}{A} \right)^{1/n} \frac{1}{n} I^{(1/n-1)} = \frac{1}{n} \frac{E}{I}, \quad (3)$$

where E/I is the dc resistance at the particular current at which n also is evaluated.

Thus, we see that the ac resistance of the varistor is the dc resistance divided by the exponent n , and further, that the ac resistance is con-

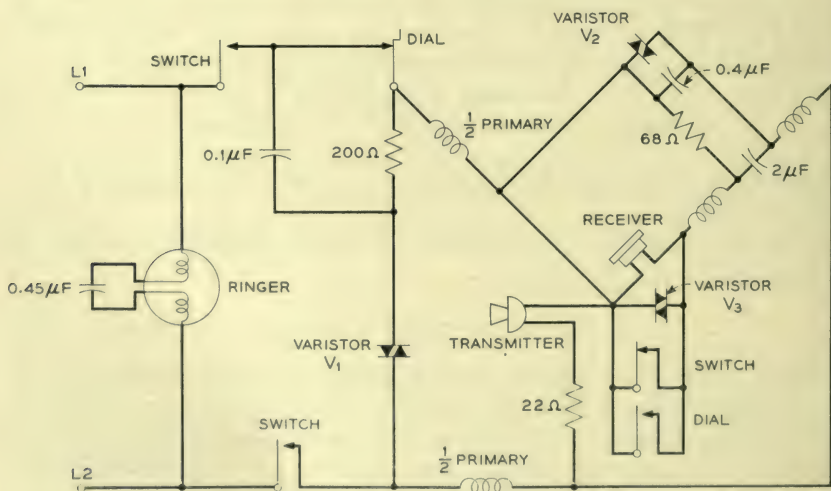


Fig. 2 — Circuit schematic of 500D telephone set.

trolled by the value of direct current. This is of great significance in the control of sidetone and transmission equalization, as will be apparent later.

Silicon carbide varistors have been used as circuit elements for various purposes for a number of years. In fact, the V_1 varistor was used in the original design of 500-type set to protect the tungsten filament against abnormal voltages. In this instance it was shunted directly across the filament, so that its inherent property of decreasing in resistance with applied voltage would serve to bypass current from the filament and thereby prevent damage when the applied voltage became too high. The requirements imposed by the transmission equalization and sidetone balance functions for the 500-type set are quite severe. A new varistor, the V_2 , had to be developed which not only was much lower in resistance, but also maintained the n of the voltage-current characteristic at a high value. The successful outcome of this development and of the accompanying problems of design and manufacture, were of critical importance in making possible the simplification and economies offered by the 500D set.

TRANSMISSION EQUALIZATION

Fig. 2 is a circuit schematic of the 500D telephone set. This circuit is a modification of the circuit used in the 500B telephone set, which in turn is one of the Campbell anti-sidetone circuits² that has been standard in the Bell System for a number of years.

A simplified diagram of the transmission portion of the circuit is shown in Fig. 3, and it can be seen that there are three branches in the

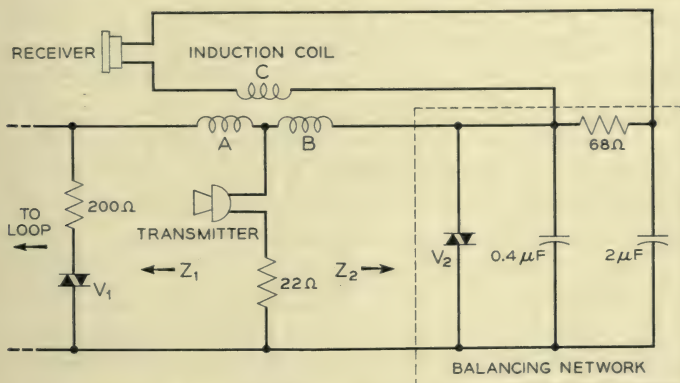


Fig. 3 — Diagram of transmission circuit of 500D telephone set illustrating side tone balance.

circuit which carry direct current. One is through the combination of the varistor V_1 and the 200-ohm resistance, which is bridged across the set. Another is through the winding A of the coil, the transmitter and 22-ohm resistance. The third is through winding B of the coil and the varistor V_2 .

Fig. 4 shows the dc and ac resistance of the V_1 and V_2 varistors at various loop currents when connected in the telephone circuit in the normal way. The magnitude of the direct currents which flow through the varistors under these conditions differs from the currents shown in Fig. 1 and, therefore, the resistances are different. However, the inherent property of decreasing in resistance with increasing current is evident, as is also the fact that the ac resistance is much lower than the dc resistance. From this it is apparent that the varistors will introduce shunt losses and that these losses will be greater for ac than for direct currents because the resistance of the varistor is lower in the former case.

The variation in transmitting plus receiving loss with loop current is shown in Fig. 5. These curves are plotted in terms of the level of the 302-type set. If no equalization were provided in the circuit, the full transmitting and receiving gain offered by the new transmitter and receiver would be obtained at all loop currents. This condition is indicated by the broken line at the top of the diagram. With the equalization introduced by the varistors, the loss is graded from an insignificant

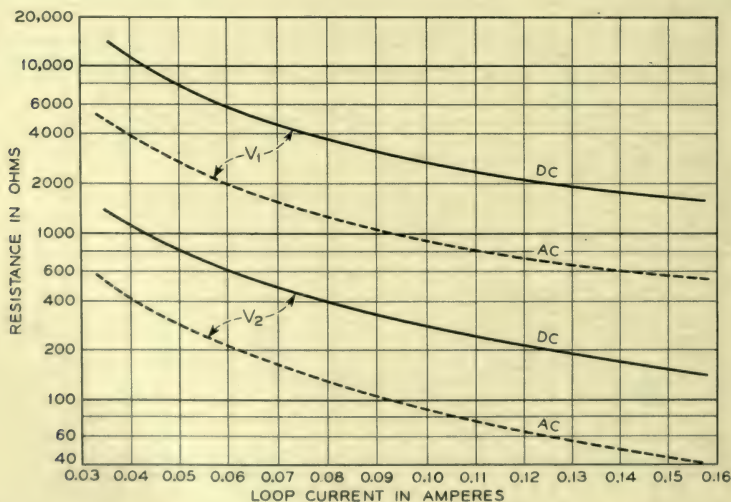


Fig. 4 — Characteristics of V_1 and V_2 varistors in circuit of 500D telephone set.

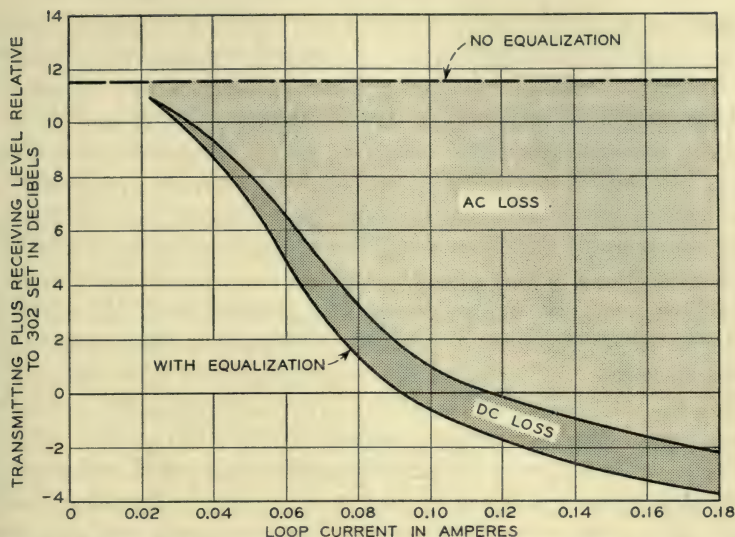


Fig. 5 — Level of 500D telephone set at various loop currents with and without equalization.

amount for long loops to a maximum for short loops. It is important to note that the greatest portion of this loss is the ac shunting loss introduced by the varistors.

SIDETONE BALANCE

Before considering the matter of sidetone balance, it will be helpful to examine some of the features of the telephone set circuit which are required for good receiving and transmitting performance. For simplicity, the induction coil is shown in a hybrid arrangement in Fig. 3. The speech currents received from the loop pass through the windings A and B of the induction coil which are so connected that they produce additive voltages in winding C which is connected to the receiver. These additive voltages would cause a resultant current to flow in the network resistance (68 ohms), were they not opposed by an approximately equal voltage 180° out of phase which results from the receiving current in winding B. Therefore, there is little or no voltage drop across the network resistance. Maximum receiving levels are thus obtained without appreciable power loss in the network resistance.

The transmitter of the set is made low in resistance because it is in series with the loop and influences the permissible loop range of the set from the standpoint of supervision. The impedance Z_1 looking toward

the loop is relatively high. To obtain acceptable transmitting levels, the low impedance of the transmitter must be approximately matched to the high impedance loop by providing a proper ratio of turns between windings A and B. Therefore, the impedance Z_2 looking toward the balancing network must be made low and since an important element of this network is the V_2 varistor, this had to have an unusually low resistance. This is the reason why it was necessary to undertake a special development to make available a silicon carbide varistor having suitable characteristics.

Sidetone would be eliminated if the vectorial sum of the voltages in the mesh which includes the receiver, and which arise as a result of speech and noise picked up by the transmitter, were zero. The complete elimination of sidetone is not desirable, but the objective is to keep it at a low level by a balance of opposing voltages. To achieve this result any voltage developed in the local transmitter is divided in the windings A and B so that the voltages induced in winding C are opposing. Furthermore, the voltage across the network resistance arising from current flowing in winding B is arranged to oppose this resultant of voltages induced in winding C. The overall effect of this balance is that the current in the receiver as a result of voltages developed in the transmitter is small. Also, this result gives maximum transmitting levels because there is little power loss in the receiver. However, to maintain the balance which gives low receiver currents the impedances Z_1 and Z_2 , as affected by the transformation ratio of the coil, must be comparable. This is the key to good sidetone balance — that the impedances Z_1 and Z_2 be effectively matched both in magnitude and phase.

Now in the telephone plant, the impedance looking from the station toward the loop varies widely from one installation to another, and even from one call to another. The loop may be long or short, of small gauge or large gauge, and composed of cable or open wire or combinations of the two. Furthermore, the loop may be loaded or non-loaded and it may be connected through central office circuits of different characteristics to a trunk or distant loop and telephone set which also may cover a range in impedance. It is quite obvious that under these conditions the impedance looking toward the loop will not only vary over a wide range in magnitude, but may be inductive, or capacitive or essentially resistive.

If we assume that the V_2 varistor were not present, the impedance Z_2 looking toward the balancing network is not influenced by loop current. This is the situation that has obtained in balancing sidetone in preceding designs of telephone set. One of the impedances to be matched

was fixed and the other varied over a wide range. The sidetone balance under such conditions represented the best compromise that could be made among a large number of uncontrolled factors.

Let us examine next how the varistors affect this balancing problem and consider the influence of long and short loops. From what has already been shown, the varistor is a variable impedance element — that is, both its dc and ac resistance depend on the direct current. If the loop is long and composed of open wire the impedance looking toward it is high, perhaps 1200 ohms, and the direct current is low. The dc resistance of the V_1 varistor is then so high (approximately 10,000 ohms) that it has no appreciable effect on the impedance Z_1 . Even the ac resistance of about 4,000 ohms does not have an appreciable shunting effect on Z_1 . On short loops and local connections where the current is large and the impedance looking toward the loop may be of the order of 900 ohms, the dc resistance of 1,800 ohms has little effect. However, the ac resistance of 600 ohms does have an appreciable effect on Z_1 , bringing about a better impedance match.

The impedance looking toward the balancing network and receiver without V_2 connected in the circuit is approximately 75 ohms, and is not affected by the loop current. If we go through the same process of examining the shunting effect of V_2 on the impedance Z_2 , we see that at small loop currents both the dc and ac resistance of V_2 are too high to have any appreciable effect. At large loop currents the dc resistance of V_2 is under 150 ohms, but the ac resistance has dropped to well below 50 ohms and has a decided effect on the impedance.

Since the varistors are essentially pure resistances they tend to make the impedances Z_1 and Z_2 presented to the coil more resistive than they would be if the varistors were not present. This has the effect of reducing the difference between loop impedances which are capacitative, inductive, or resistive from the standpoint of phase angle and thereby improves the impedance match.

The overall effect of the automatic changing in resistance of the V_1 and V_2 varistors is to provide better matching and thereby reduced sidetone. This is shown in Fig. 6, where the loudness level of sidetone of the 500D set is plotted against loop current. It will be noted that with V_1 and V_2 open circuited, there is a continuing increase in sidetone loudness level with increasing loop current. When both V_1 and V_2 are connected in the circuit a decrease in sidetone level of 10 db or more is obtained at medium and large loop currents. The decrease is less at small loop currents. The close interrelation between the varistors and the circuit is well illustrated by the fact that if either the V_1 or V_2

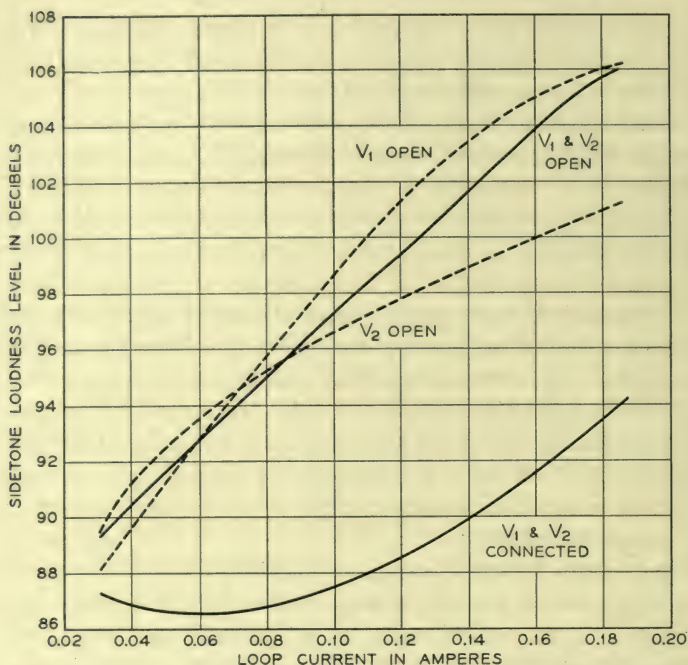


FIG. 6. — Effect of V_1 and V_2 varistors on sidetone loudness levels of 500D telephone set for average speech.

varistor is opened, almost all the effects of sidetone balancing are lost. It requires the presence of both V_1 and V_2 to obtain the impedance matching necessary for good balance.

From all that has been said, it should not be thought that the sidetone balance of the 500D set is perfect under all plant conditions. The impedance looking toward the loop varies so greatly in magnitude and phase that the best overall balance involves compromise. However, the balance provided is as good as with the original design of 500 set, which was considerably better than that attained by any previous design of commercial telephone set.

OTHER MODIFICATIONS

The introduction of the new equalization and sidetone controls in the telephone set has required a number of changes in other components of the set to realize all possible economies. Fig. 7 is a photograph of the transmission network (425A) which was employed in the earlier 500 A and B type sets. In the new arrangement shown in Figure 8



Fig. 7 — 425A Network used with early design of set — the 500A and B type.

the V_1 and V_2 varistors have been included in the network (425B), thereby providing mechanical protection for these elements. It has been necessary to design a new terminal block which is the plastic member to which the components of the network are mounted.

OVERALL RESULTS

By means of the V_1 and V_2 silicon carbide varistors and the rearrangements made in the circuit, the overall transmission performance of the modified set has been made substantially the same as that provided by the original design of set. The savings in manufacturing cost resulting from this modification are significant. Figure 9 shows on the left the

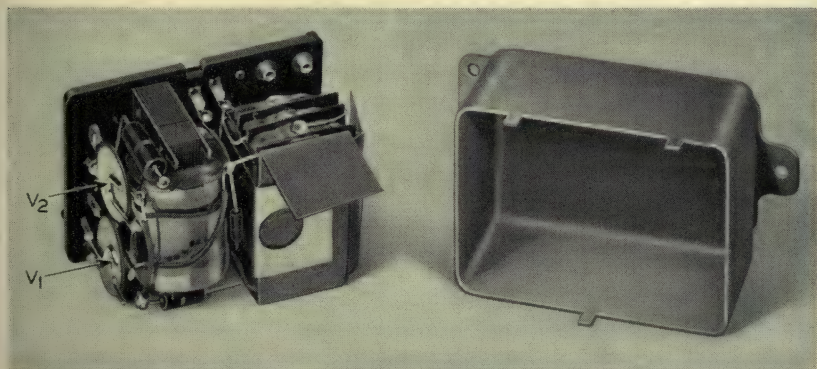


Fig. 8 — 425B Network used in present production set — the 500C and D type, showing the V_1 and V_2 varistors.

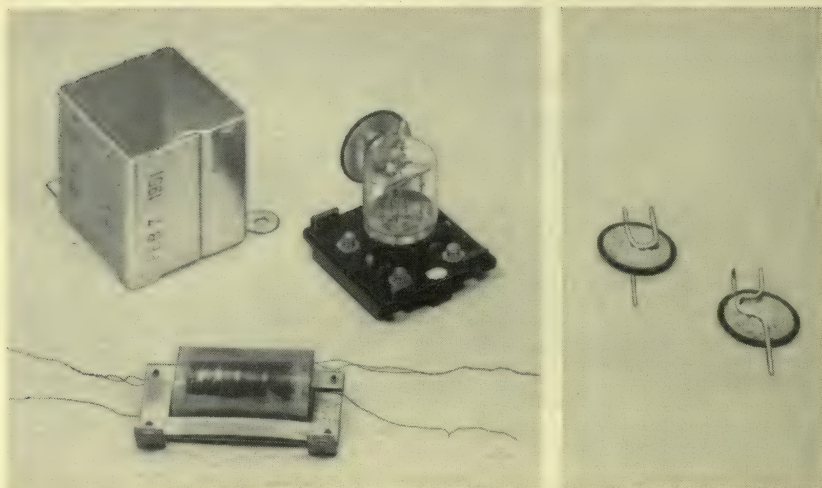


Fig. 9 — The 311A equalizer and autotransformer on the left are replaced in the 500C and D set by the V_1 and V_2 varistors on the right.

parts which have been eliminated — the 311A equalizer and auto-transformer. On the right are shown the V_1 and V_2 varistors which provide essentially the same performance as the replaced parts.

Another advantage of the change in design, though a difficult one to evaluate, is that of dependability. The silicon carbide element has an extremely long life and in this respect is inherently better than the type of equalizer which involves the use of a heated tungsten filament.

OPERATION IN NOISY LOCATIONS

Since the previous paper on the 500 set was published,¹ a study has been completed on the comparative performance of the 500-type set and its predecessor, the 302 type, in locations where room noise is severe. This is of particular interest, because the sidetone characteristics which have been discussed play an important part in transmission performance under such conditions.

For many years the securing of improved performance of the telephone set under noisy conditions has been a serious problem. It has been established that receiving impairment in a noisy location is far more limiting than the transmitting impairment resulting from the noise picked up at the noisy location and delivered to the distant party. This is due to the fact that the signal to noise ratio in the transmitting direction is improved by the natural tendency of the speaker to raise

his voice when the ambient noise level is high. A major factor in the impairment of received speech in the noisy location is the masking of received speech by noise picked up by the transmitter and delivered via the sidetone path to the receiver. Therefore, a telephone set which minimizes the masking sidetone noise during the listening interval should operate more satisfactorily in noisy locations. In many previous instances where the noise conditions were extreme, a "push-to-listen" switch was provided to cut off the local transmitter while listening and thus prevent the introduction of noise to the receiver via the sidetone path. With its higher efficiency and better sidetone balance, the 500-type set approaches the good performance of such a "variable" set under noisy conditions and provides this adequate performance to the user more conveniently and with the expenditure of less effort.

Laboratory tests have been made which permit appraisal of the merits of various telephone sets or modifications of them under high noise level conditions. These include talking and listening tests between two telephone stations under typical loop and central office conditions with an adjustable amount of typical room noise at the listening end. During the tests the length of trunk between the two stations was increased until the received speech was just sufficiently intelligible to permit carrying on a conversation. This threshold of intelligibility expressed in db of trunk loss was taken as a criterion of the performance of a given set.

Fig. 10 shows the results of the tests made with the 500-type set compared with the 302-type set, and with variations of the 500 type set. 500C- and D-type sets having silicon carbide varistors were used in this comparison, but in this respect the performance is no different than for the earlier design (500A and B). The curves shown are for the average of eight observers.

The abscissae of the curves are the noise levels at the listening end expressed in db above 10^{-16} watts per sq. cm. For reference purposes it should be kept in mind that the average noise in a fairly large business office is, on this scale, approximately 65 db. For further orientation as to the significance of the noise levels shown, it should be noted that when a level of about 100 db or higher is reached, it is extremely difficult to carry on a face-to-face conversation where no telephone is involved. The ordinates show the relative trunk loss in db over which it is possible, under the stated conditions, to just carry on a telephone conversation. Therefore, the larger the trunk loss, the better the performance.

It is seen from Figure 10 that the 500-type set is from 6 to 8 db better than the 302 set over the indicated range of ambient noise conditions.

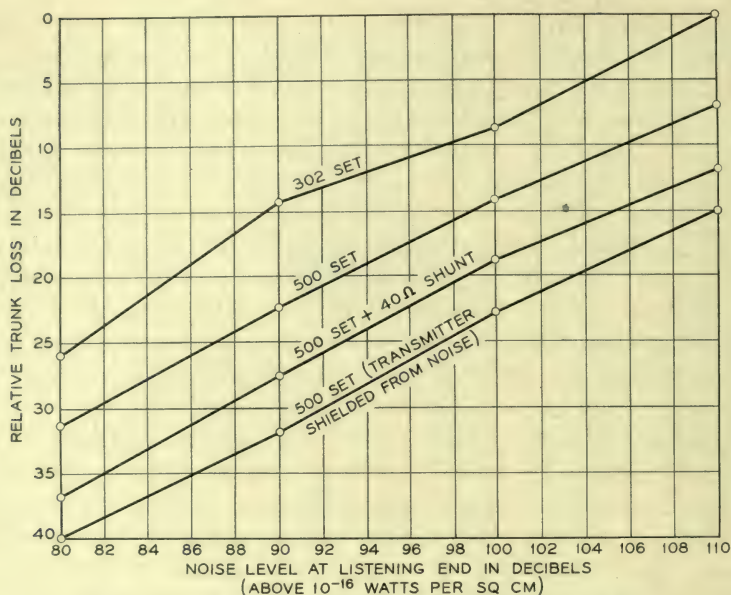


Fig. 10 — Effect of noise at listening end on trunk loss.

This improvement is attributed to the following:

1. The better sidetone balance of the 500-set circuit reduces the level of sidetone noise picked up by the transmitter and reproduced by the receiver, thereby providing a gain in signal to noise ratio.

2. The acoustic impedance looking from the ear canal into the U1 receiver³ of the 500-type set is lower than the acoustic impedance of the HA1 receiver of the 302 set and, therefore, the acoustic noise pressure built up in the ear canal by leakage around the receiver cap is lower than when the HA1 receiver is used.

3. The output from the transmitter of the 500-type set under noise conditions contains less extraneous distortion products, giving it a character which is less disturbing than with the 302 set.

Included in the laboratory tests were experiments in which the output of the transmitter was intentionally lowered. It is known that a speaker in a noisy location instinctively raises his talking level. Since it is desirable to keep the sidetone level low to improve the signal to noise ratio, the output level of the transmitter at the listening end was reduced about 10 db by shunting a resistance of approximately 40 ohms across the transmitter. The sensitivity of the transmitter is reduced

and sidetone noise is directly lowered. The level of the received speech is unaffected. Therefore, the signal to noise ratio in the receiving direction is improved. The transmitting level is not too adversely affected, because the tendency of the subscriber to increase his talking level when the room noise level is high largely offsets the lower efficiency of the transmitter to speech sounds. These effects are borne out in the applicable curve of Fig. 10 which shows that an improvement of approximately 5 db in trunk results from the shunting of the transmitter.

Carrying this experiment further and acoustically shielding the local transmitter completely from noise, the lower curve of Fig. 10 was obtained. This is equivalent to shorting the transmitter as is done in the earlier "push-to-listen" types of telephone set, except that it does not introduce any electrical effects in the circuit. This shielding provides about 4 db additional discrimination against noise. This, then, indicates the limit to which we can go in reducing the effects of room noise on received speech by operating on the transmitter element alone.

From the data presented in Fig. 10 it is evident that the 500-type telephone set provides a significant improvement when the subscriber is carrying on a telephone conversation under noisy room conditions. While it has been indicated above that the greatest gain of the 500-type set over the 302 type under high room noise conditions is with respect to received transmission, it is also better in transmitting from noisy locations. This is because of the less disturbing character of the transmitted noise and because the signal to noise ratio is improved by having the transmitter closer to the subscriber's lips, which is a feature of the new handset.

Where noise conditions in the field are severe, the 500-type set will provide material improvement. Typical of such conditions are noisy business locations, telephone stations which are located in power plants and near loading ramps at airports. Where the noise is particularly severe, the provision of a 40-ohm shunt resistance on the transmitter of the 500-type set offers still further improvement and is recommended for application by the local telephone people.

CONCLUSION

The simplification of the 500 set made possible by the use of silicon carbide varistors, with its attendant reduction in manufacturing and maintenance costs and increase in life represents a significant step forward in station set design. A valuable characteristic of the 500-type set is the substantial improvement in transmission performance under conditions of severe ambient noise.

REFERENCES

1. Inglis, A. H., and W. L. Tuffnell, An Improved Telephone Set. Bell System Tech. J. **30**, pp. 239-270, April, 1951.
2. Gibbon, C. O., The Common Battery Anti-Sidetone Subscriber Set. Bell System Tech. J. **17**, pp. 245-257, April, 1938.
3. Mott, E. E., and R. C. Miner, The Ring Armature Telephone Receiver. Bell System Tech. J. **30**, pp. 110-140, Jan., 1951.
4. Inglis, A. H., Transmission Features of the New Telephone Sets. Bell System Tech. J. V. **17**, pp. 358-380, July, 1938.

Automatic Line Insulation Test Equipment for Local Crossbar Systems

By R. W. BURNS and J. W. DEHN

(Manuscript received February 9, 1953)

Moisture entering faults in the insulation of subscriber lines provides so-called "leakage" paths which reduce the insulation resistance. Testing the insulation resistance of all lines under the environmental conditions which tend to produce these leakages is a maintenance technique, relatively new, for detecting the insulation defects. The faults can then be corrected before they become serious enough to affect the customers' service. Subscriber reports are thereby reduced and the correction of the faults on a preventive maintenance basis tends toward a more uniform work load for the repair personnel. Rapid testing of the lines is necessary, otherwise the environmental conditions may change and the leakages will disappear without detection.

Rapid line insulation testing is practiced quite generally in all the switching systems throughout the Bell System, but the testing arrangements used are wholly or partially manually controlled in the testing and recording operations. While the benefits derived from rapid line insulation testing apply to all systems alike, this article is confined to a discussion of the entirely automatic testing and recording arrangements which are now being introduced in the No. 1 and No. 5 crossbar systems.

INTRODUCTION

The insulation resistance of subscriber lines is an important consideration in the design and operation of central office switching circuits. If the insulation resistance between the two conductors of the line, or the insulation resistance from the "ring" or "battery side" to ground, becomes low enough, the "leakage" current flowing produces the same effect as lifting the handset and failing to dial or to pass a number to the operator. This condition is called a permanent signal and the line is said to be "permanent." As long as the condition persists, the line is out of service both to incoming and outgoing traffic. Insulation resistance of a slightly

higher value will not cause a permanent signal, but its presence may interfere with other circuit functions, for example, dial pulses are distorted and a wrong number may be reached, the ringing signal may be tripped before the called party answers, or the switching circuits may fail to restore on hang-up of the receiver. If the insulation resistance is at least as high as the design limit, failures of the kind described above will not occur. The design limit for some switching systems used in the Bell System is 10,000 ohms; for others 15,000 ohms.

WHERE LINE LEAKAGES OCCUR

The outside plant distribution system for exchange lines usually consists of some underground cable, many miles of aerial cable to distribute the line conductors throughout the area and a small amount of open wire on the fringes. The insulation resistance of the cable conductors is normally quite high. If, however, a break in the cable sheath occurs moisture may enter and be absorbed by the paper insulation of the conductors near the break. This reduces the insulation resistance of the affected cable pairs. Sheath breaks may exist for a considerable length of time without reducing the insulation resistance sufficiently to cause any reaction on customer service. Eventually these sheath breaks will, if not detected and repaired, permit the entrance of sufficient water during a rain to reduce the insulation resistance to the point where permanent signals occur on several pairs. Then repairs must be made on an emergency basis to guard against a complete failure of all line circuits in the cable. One of the common causes of sheath breaks in some residential areas is gnawing by squirrels — squirrel bites.

Cable terminals are located on the poles or on the cable for making drop wire connections to the customers' premises. Binding posts mounted in a face plate within the terminal are connected to some of the cable pairs through a terminal cable stub joined to the aerial cable. Each cable pair is thus connected to binding posts in about four or five terminals on the average. When water or condensation wets the face plate, leakage currents will flow between the binding posts. If dirt and dust have accumulated on the face plates, the combined resistance of all leakage paths in parallel across the pair or to ground may become relatively low.

While open wire makes up only a small part of the outside plant circuits most areas have some lines containing from a few spans up to several miles of open-wire conductors. It is difficult to keep the open-wire plant in a condition so that it will be free from leakages under wet

weather conditions. Trees grow up into the wires and during rainy weather the branches often drop across the wires. When this occurs at many points on the open-wire run the combined leakage along the pair may become very low. In some localities, moss growing on the wires, salt spray or heavy fog conditions causing leakage at insulators are additional causes of low insulation resistance on open-wire lines.

Drop wire used to make the connection from the cable terminal to the customer's premises is subject to damage by abrasion and the insulation deteriorates from the effects of the weather. The old style of drop wire, a large amount of which is still in use in the plant, is insulated with a rubber compound and covered with a water proof cotton braid. When this braid protection is lost due to abrasion or effects of the weather after many years in service cracks form in the rubber insulation. Moisture enters these cracks in wet weather causing low resistance leakages. It is expected that the latest type of drop wire with the tougher neoprene covering will withstand the hazards for a longer period of time than does the old drop wire but undoubtedly the end of its useful life will ultimately be determinable by measurement of the insulation resistance under wet weather conditions.

Inside wiring on the customer's premises often remains in service for a long period of time and the insulation deteriorates over the years. If the insulation is in poor condition, inside wiring will develop leakages during periods of prolonged high humidity indoors which occur frequently during the summer months.

EFFECTS OF WEATHER ON LINE INSULATION RESISTANCE

When weather conditions are favorable — clear weather with low humidity — the insulation resistance of the line conductors in the outside plant is quite high compared to the design limit for central office switching circuits. When the plant becomes thoroughly wet from a hard rain the insulation resistance drops very considerably because of leakages which are in parallel all along the line. Fig. 1 shows this very clearly. The data for the curves in this figure were collected in a special study conducted in 1931 to determine the insulation resistance distribution of exchange lines in underground and aerial cable and open-wire plant under different weather conditions — dry, humid and wet. About 6,000 dial lines selected at random in twelve large cities in different parts of the United States were tested under different weather conditions and no repairs were made throughout the study period except to correct unsatisfactory service conditions. The tests were made with the voltmeter

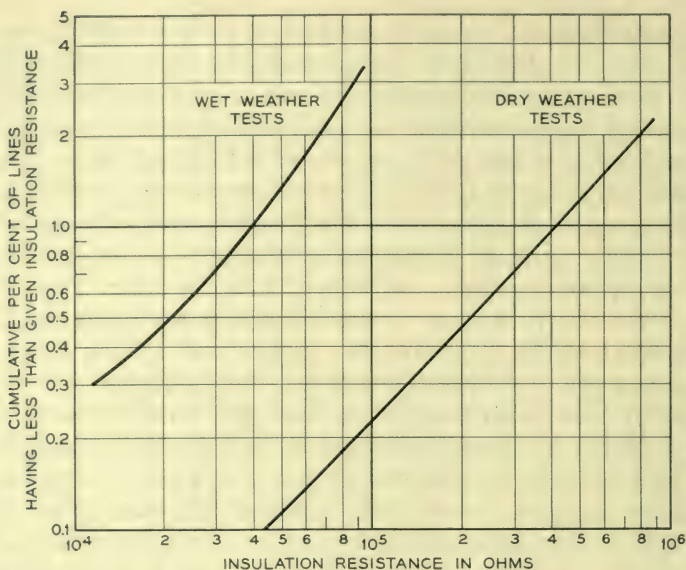


Fig. 1 — Wet weather versus dry weather tests — all types of Outside Plant combined.

test circuit of the local test desk which is used for testing subscriber lines reported in trouble. The speed at which lines can be tested from the test desk is necessarily slow because each number must be dialed or called individually therefore, only a small sample of lines was selected from any one office. Rapid line insulation test equipment was not in use in any area when this study was made.

The upper curve of Fig. 1 shows that about 0.3 per cent of the lines were below 11,000 ohms in wet weather while in dry weather only about 0.3 per cent of these same lines were below 140,000 ohms. Similarly, in wet weather two and one-quarter per cent of the lines were below 47,000 ohms but in dry weather only two and one-quarter per cent were below 900,000 ohms. The same wet weather test data summarized by types of outside plant is shown in the curves of Fig. 2. Lines in underground cable only, show the highest line insulation resistance. Those with aerial cable are next and those with some open wire have the lowest insulation resistance.

Fig. 3 shows the wet weather line insulation resistance distribution of 11,600 lines in an eastern city where the exchange outside plant had been thoroughly reconditioned prior to conversion from manual to cross-bar dial operation in 1940. The upper curve represents the approximate distribution before reconditioning and the lower curve represents the

distribution after the reconditioning was completed. The lower curve can be considered as representative of an outside plant in good condition as of the year 1940 when rapid line insulation testing was not yet used. Since that time, a considerable number of improvements leading to better insulation resistance characteristics have been made in outside plant items, such as drop wire and cable terminals, and a higher average insulation resistance would currently obtain. However, during wet weather, an undesirably large number of lines would still be closer than desired to the design limit.

MAINTENANCE WITHOUT RAPID LINE INSULATION TESTING

It may appear at first sight that the per cent of lines near or below the design limit is so small as to be unworthy of particular notice. However, an examination of this from a maintenance standpoint will prove otherwise. The testing of lines reported in trouble and dispatching of craftsmen on the outside to make repairs are handled from a local test center which on the average serves about 50,000 stations. There will be about six local test desk positions manned to do the testing and dis-

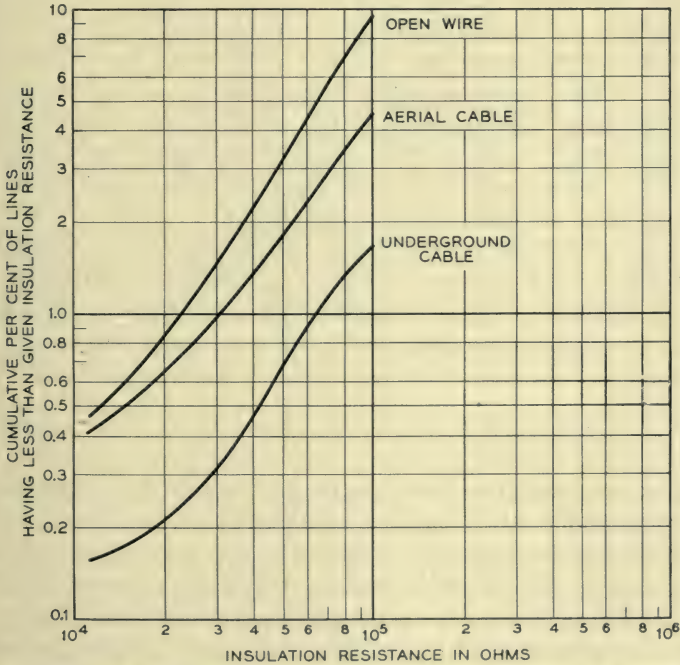


Fig. 2 — Wet weather tests — insulation resistance by types of Outside Plant.

patching work. Such a test center would handle at the current trouble rate about 120 reports on subscriber lines per day. If the outside plant is in a condition represented by the lower curve of Fig. 3 where 0.2 per cent of the lines are below 10,000 ohm resistance, this represents 100 additional stations in the region where service reactions may be expected from wet weather conditions. Consequently where these plant conditions obtain, there is usually a noticeable increase in subscriber's reports in wet weather and the repair load rises sharply. Without line insulation test equipment to test the lines rapidly while the low insulation resistance obtains, there is no way to pick out the worst lines from an insulation resistance standpoint after the weather has cleared. Visual inspections are costly and superficial indications do not always give reliable evidence of low insulation resistance. Consequently the large percentage of repairs to correct insulation defects are made as a result of subscriber reports and only a small per cent by routine preventive maintenance where rapid line-insulation testing equipment is not used.

MAINTENANCE WITH RAPID LINE INSULATION TESTING

The only experience to date with the fully automatic line insulation test equipment for crossbar offices is in the Media, Pa., No. 5 crossbar office. This test equipment has been in use for about one year and low insulation resistance cases recorded on each test have been investigated and repairs made. The condition of the outside plant in the exchange area is represented in Fig. 4. The curves in this figure are based upon the results of tests of the 4200 working lines in the summer of 1952 under very wet conditions of the outside plant.

The small percentage of lines below 40,000 ohms indicates that the poor insulation cases are being corrected well before reaching the stage where the customers' service would be affected. This is done with a minimum of maintenance effort as the test equipment spots the line automatically.

DETECTING SHEATH BREAKS

While sheath breaks in aerial cable are brought to light by rainy weather it is inadvisable to wait for rain to disclose them because of the possible serious effects on service and the need then for repairs on an emergency basis. Rapid line-insulation testing techniques have been very successful in disclosing sheath breaks in aerial cable. During the night the cable sheath cools and as the pressure inside decreases, outside air with a high moisture content enters the sheath break. The paper

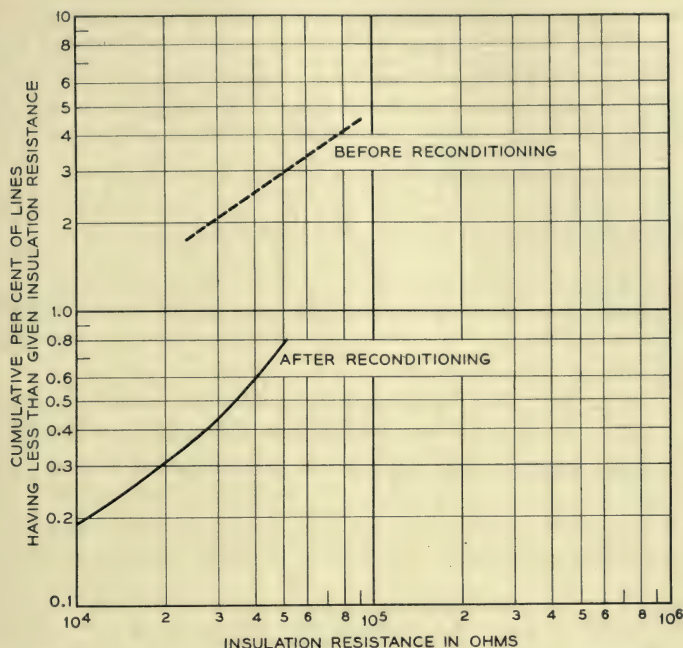


Fig. 3 — Wet weather tests — reconditioned Outside Plant.

insulation of one or more pairs near the break absorbs the moisture and the insulation resistance of the affected pairs is lowered. Tests made from about 3 A.M. to 5 A.M. are effective in detecting these leaks and by applying the latest fault locating methods, many of the sheath breaks can be located. The test equipment is so arranged that, for the most part, lower leak conditions on the line which may be present outside of the aerial cable will not register on these tests for cable defects.

BENEFITS DERIVED FROM RAPID LINE-INSULATION TESTING

With the aid of rapid line-insulation test equipment the maintenance personnel can correct the greater part of insulation defects on a preventive maintenance basis which results in a reduction of subscriber reports. Service to the customer is thereby improved, maintenance effort is reduced and repairs on an emergency basis often involving the expenditure of overtime are required less frequently. Rapid line-insulation testing is also of great value in rapidly determining the extent of storm or flood damage and makes it possible to direct immediate efforts where the greatest benefits will be derived.

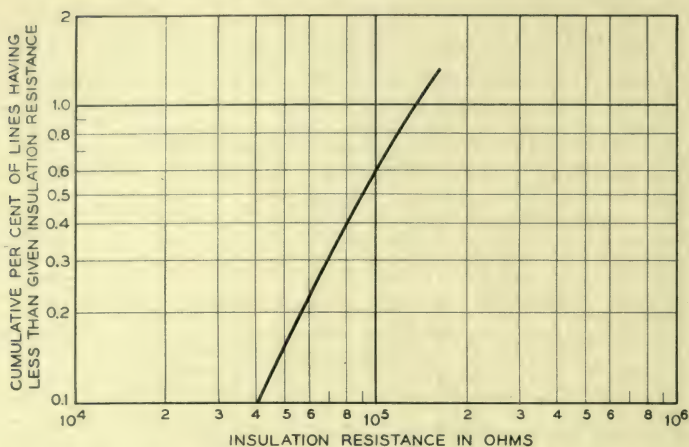


Fig. 4 — Wet weather tests — automatic line insulation testing equipment used.

CROSSBAR TEST EQUIPMENT — TYPES OF TESTS AND SENSITIVITIES

The crossbar line-insulation test equipment is arranged to make three different kinds of tests from the standpoint of the way in which the resistance measuring circuit (line-insulation test circuit) is connected to the line and to battery or ground. Each kind of test may be made in three different resistance ranges, hence there are nine different tests. When the test equipment is performing one of these tests a record is made of all lines which have an insulation resistance below the top limit of the particular test. Each test range is divided into three bands of resistance and the record indicates whether the insulation resistance lies within the first quarter (low band), the second quarter (medium band) or the upper half (high band) so that preference can be given to the worst cases in clearing the trouble. The test numbers, ranges and resistance bands are shown in the Table I.

SHORT AND RING GROUND TEST

The first kind of test is called "short and ring ground test." This test measures the leak in the way in which it affects the line circuits, pulsing circuits and supervisory circuits of the central office switching system. The line-insulation test circuit (LIT circuit) which is described later on is connected to central office battery potential and to the line in the manner shown in Fig. 5 (a). As indicated in this illustration, leakage resistance from ring to tip and from ring to ground are measured

TABLE I — CROSSBAR LINE-INSULATION TESTS

Types of Tests and Test Numbers			Range	Resistance Bands* (kilo-ohms)		
Short and Ring Ground	Tip and Ring Ground	Foreign E.M.F.		Low	Medium	High
1	4	Not used	A	0-40	40-80	80-160
2	5	7	B	0-160	160-320	320-640
3	6	8	C	0-640	640-1250	1250-2500
Not used	Not used	9	D	0-1250	1250-2500	2500-5000

* The equipment is arranged so that by cross connection changes the bands of resistances in each range may be halved or doubled, if necessary, to meet local conditions.

in parallel. If the line-insulation resistance is more than the top limit of the test range, the test equipment proceeds to the next line. If, however, the resistance measured is less than the top limit the test equipment immediately retests the line with the central office ground removed as shown in Fig. 5 (b). This re-test measures the leakage from ring to tip. If leakage from ring to tip only is indicated the drop wire is probably the cause.

The No. 1 test is run under wet weather conditions to detect all lines which are below the top limit of this test (normally 160,000 ohms). If the maintenance force keeps these cases cleared out by running tests during every rain the wire plant can be kept in good condition. During long periods of dry weather all line insulation test indications previously recorded may have been investigated, then if a light rain occurs it may be desirable to run the No. 2 test to provide a satisfactory number of failure indications for subsequent investigation. The No. 2 test can also be used to detect leakages in inside wiring which occur frequently during the summer when houses are unheated and the inside humidity

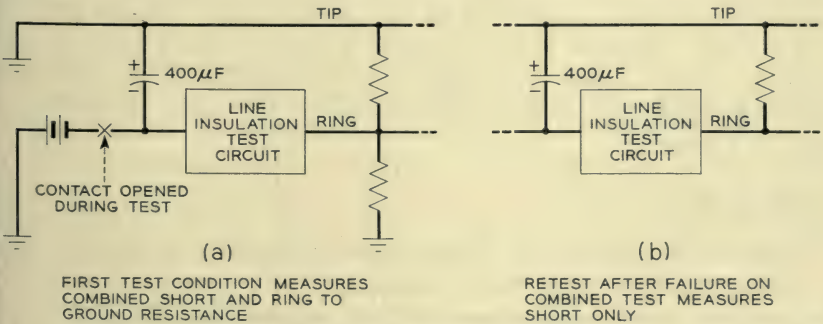


Fig. 5 — Short and ring ground test.

is very high. Tests to disclose insulation defects in inside wiring are made when the weather is clear with low outside humidity at which time leakages in other parts of the exchange plant will rarely be found.

TIP AND RING GROUND TEST

The "tip and ring ground" test can be made in three ranges of sensitivity as shown in Table I. The arrangement of the test circuit for this type of test is shown in Figure 6. The first test condition, Fig. 6 (a), measures the combined leakage resistance from tip and ring to ground. The tip and ring are connected together therefore leakage across the pair is not indicated. This eliminates practically all of the drop wire

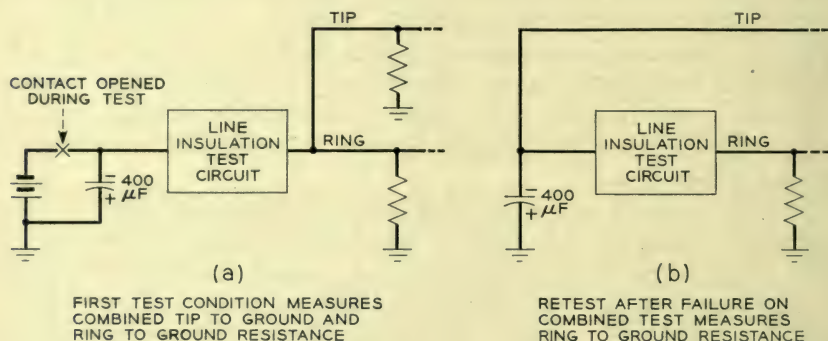


Fig. 6 — Tip and ring ground test.

leakages. If a failure is indicated on the first test condition, a re-test is made under the condition shown in Fig. 6 (b) to check only the resistance from ring to ground. If the ring tests clear it is known that the leakage is from tip to ground. This kind of test is of particular value in checking terminal face plate leakages which are predominately leakages from tip to ground. This type of test can also be used to check that tip conductors on party lines in message rate areas are free of low resistance grounds which might result in wrong party identification on calls made by the ring party.

FEMF TEST

The FEMF (foreign e.m.f.) test is used to measure leakages in cable to detect sheath breaks. These leakages are high resistance compared to other leaks which may be present across the pair connected for test or from the pair to ground. To make the latter ineffective so as not to

mask the higher resistance cable leakages the test circuit is grounded and connected to the line with the tip and ring short circuited as shown in Fig. 7. The test circuit measures the leakage current flowing from the pair connected for test to the ring conductors of other subscriber lines which are connected to battery potential in the central office. Leakages outside the cable will not cause leakage currents to flow through the line insulation test circuit. This test is called FEMF because it measures leak in the same way as does the test desk voltmeter when the FEMF test key is operated to connect ground instead of test battery to the voltmeter. Leakages as high as about 2 megohms can be successfully located after the cable has been identified by analysis of pairs affected by leaks.

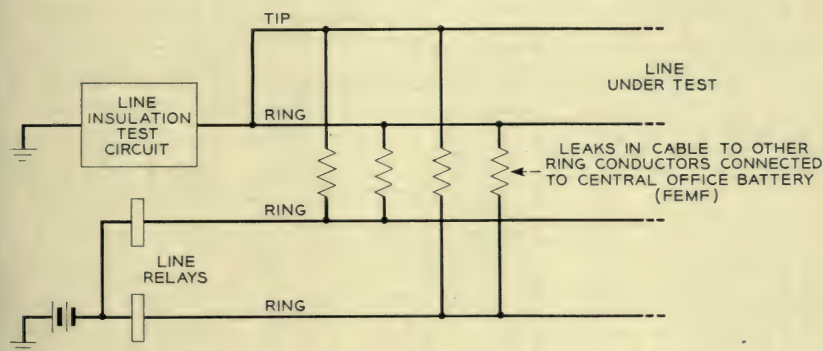


Fig. 7 — Foreign E.M.F. test.

TEST AND RECORDING CIRCUITS

The equipment for automatically measuring and recording line-insulation resistance consists of three principal parts. First, the device for measuring the low currents involved; second, the means for connecting this measuring device to each of the lines in rapid succession; and third, the apparatus required for recording the faults discovered.

The first of these, known as the line-insulation test circuit, is capable of detecting insulation resistance as high as ten megohms, with fast enough response, so that a satisfactory rate of line testing (about 10,000 to 12,000 lines per hour) is attainable. It is provided with a filter to attenuate both 25- and 60-cycle interference, so that leakage faults may be separated from other types such as induction from power lines. As shown in Table I, its sensitivity may be easily changed, so as to be compatible with atmospheric conditions, the kind of test to be made, and the condition of the outside plant. When a particular sensitivity or

"range" is chosen, the insulation fault detector will report all lines having lower insulation resistance than the chosen value, and as already stated it will further identify those lines having insulation resistance lower than one-half and one-quarter of this resistance.

This test circuit is connected to each of the lines in rapid succession by the line-insulation test control circuit. As shown in the block diagram of Fig. 8, the control circuit appropriates the control and testing wires between one of the markers (usually marker No. 2) and each of the line link frames. The arrangement shown is for a No. 5 crossbar office, but the No. 1 crossbar arrangement is similar in principle. By means of these appropriated connections the control circuit is able to select the

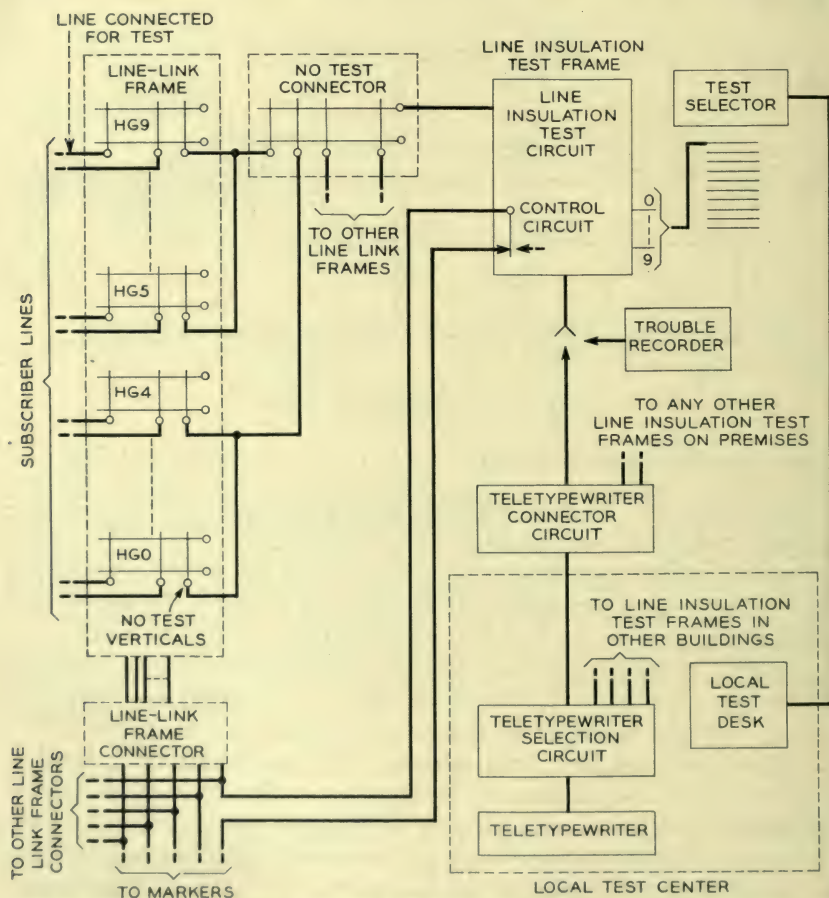


Fig. 8 — Block diagram of line insulation test control circuit connections.

lines and if the selected line is not busy its tip and ring conductors are connected to the insulation resistance measuring circuit (line-insulation test circuit). To establish this connection, the control circuit chooses an idle line link to connect the line vertical to the no test vertical and the test path is completed through the no test connector to the line-insulation test frame.

The operation of the control circuit may be started in the central office by operation of keys at the test frame, but it is more commonly started by remote control from the local test center. This permits insulation tests to be made even though the central office is unattended. When the control circuit is started, it must also be directed to make one of the three types of tests, and to choose one of the three test sensitivities for each of these types as shown in Table I. This is accomplished by operating one of nine keys at the test frame or by selecting the test trunk at the test center and dialing one of nine codes and then operating a key at the test desk. Either action chooses the type of test and the sensitivity, and causes the test circuit to start. A tenth key or a tenth code is used to stop the test before the end of a complete cycle, when this is required, so that the type of test or the sensitivity may be changed.

A pre-arranged regular pattern is followed in testing the lines. A line link frame is selected, a particular five lines are tested, and then the frame is released, and the next frame is selected, and another five lines are tested, and so on. Lines found busy, dial PBX trunks and line link frame terminations of intertoll trunks and similar circuits which would cause false operation of the test circuit are passed by without test. The line link frames are selected in order, beginning with the lowest numbered frames and continuing to the highest. When one cycle through the frames has been completed, another cycle will be started. However, on the next cycle, different groups of five lines will be tested. On the first cycle, the five lines in vertical group 0, horizontal group 0 will be tested, on the second cycle, the five lines of vertical group 1, horizontal group 0 will be tested, in each line link frame and so on until all lines in horizontal group 0 of all line link frames have been tested. On subsequent cycles the lines of other horizontal groups are tested in regular order.

Testing only five lines for each selection of a line link frame minimizes interference with traffic. Since one to two seconds is required for testing five lines, there will be only a slight delay to calls which require access to the frame. In addition, a heavy traffic load will cause the control circuit to stop insulation testing and restore to service the marker whose cabling it has been using. When the traffic has been reduced sufficiently, the test will be restarted automatically.

When a line fault is discovered, the control circuit makes use of either of two types of mechanism to record (a) the location of the line on the line link switches, (b) the type of test being made, and (c) a rough approximation of the insulation resistance. Since a substantial time is required to make this record, the line link frame is restored to service during this interval, to reduce interference with service. One of the above two devices is the trouble recorder, used only in the No. 5 crossbar offices. The control circuit connects itself to this machine in much the same way that a marker does, when it needs to record a trouble. Having made this connection, a card is perforated showing the line location and other data pertinent to the trouble indicated. The other device consists of teletypewriter equipment at the central office which transmits the required data to a teletypewriter page printer located at the local test center. The equipment at several central offices has common use of the same page printer at the test center and several test circuits in one building use the same teletypewriter transmitting equipment. This second recording arrangement is the more convenient of the two, since it produces the record at the place from which the activities of the outside plant repair force are directed. This arrangement is available for both No. 1 and No. 5 crossbar offices. A typical teletype record is shown in Fig. 9.

LINE-INSULATION TEST CIRCUIT

The insulation resistance measuring device is required to respond, not only to the very low current (five micro-amperes) obtainable with insulation resistance up to ten megohms, but it must also give good discrimination between resistance values in the order of 20,000 ohms. This is accomplished by desensitizing the measuring device by means of series and shunt resistors when a test using less than the maximum range is desired. A second requirement is that the measuring device be low in resistance so that the time constant of this resistance in combination with the line and filter capacity will be low enough to attain high testing speeds.

These leakage current amplitudes are so small that amplification is required in order to actuate the recording and control mechanisms of the measuring system. These small direct currents could be amplified by means of a dc amplifier. However, since it is easier to design and construct an ac amplifier of suitable stability, it is desirable to use a measuring device which has an alternating voltage output which varies with the direct current input.

A type of magnetic modulator, called a magnetor, which has these

desirable characteristics, is used as the current measuring instrument. It has a low impedance input circuit, in which the low amplitude dc leakage current flows. An alternating current of constant amplitude and frequency is supplied to separate windings of the magnetor, so that, as explained below, its output circuit delivers an alternating voltage which varies with the dc input. Fig. 10 shows the basic circuit, which operates as follows. Two identical windings, a and a', and two other identical windings, b and b', are wound on identical permalloy cores. Windings

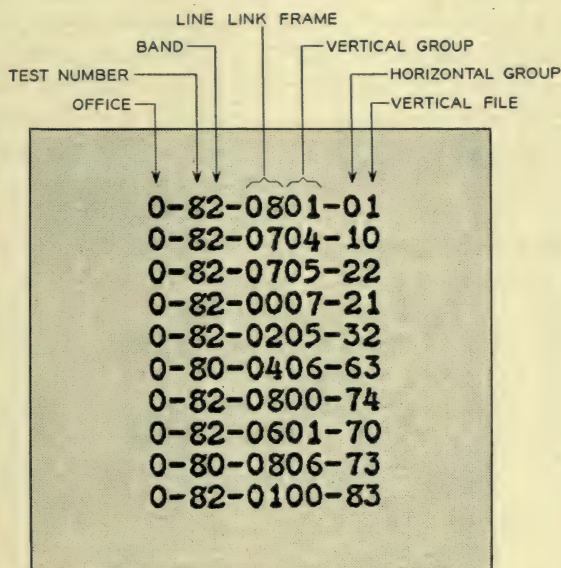


Fig. 9 — Teletype record of failures.

a and a' are connected in series and supplied with an alternating current. The magnetor cores have a characteristic as shown in Fig. 11 (b), and the amplitude of the input voltage is great enough so that the core is driven to saturation on each half cycle. Fig. 11 (c) shows the resulting flat topped flux versus time curve*. Since the voltage induced in winding b is proportional to the rate of change of flux, it will have a wave form as shown in Fig. 11 (d). The voltage peaks occur when the flux rate of change is maximum, and during the "flat" intervals the induced voltage is small. Since the b and b' windings are connected so that their output voltages are in opposition (see Fig. 11 (d)) the net output with no dc

* The wave shapes of Figs. 11 (c), 11 (d), 11 (e) and 11 (f) have been exaggerated to illustrate the action involved.

input will be zero. Manufacturing tolerances which produce dissimilarities in the windings and cores may cause a small output with zero dc input.

When the dc leakage current flows through windings b and b' , a bias flux is established in each of the cores. This causes the shape of the flux versus time curve to be changed as illustrated in Fig. 11 (e). Since the ac input will saturate the cores, both half cycles of the wave will be flat topped, but the flat portion of one-half cycle will be increased and the other decreased. Also the steeper parts of the curve will be brought closer together on one-half cycle and further apart on the other.

As illustrated in Fig. 11 (e) the dc bias will have a different effect upon the flux in each of the two cores. In one core the bias aids the positive half cycle of the input current, and in the other the negative half cycle. The result is that the core reaches saturation more quickly and, therefore the flux curve is flatter on the half cycle which is aided by the bias.

This change in the shapes of the flux time curves produces corresponding changes in the shapes of the voltage-time curves of the output windings b and b' . These are shown in Fig. 11 (f). The peaks of the voltage curves occur at the points of maximum rate of change of the flux curves, and of course the flat portions of the voltage curves (near zero) are lengthened or shortened depending upon the flatness of the flux curve tops. This skewing of the two voltage curves prevents the output voltage cancellation which was obtained with no dc, and gives

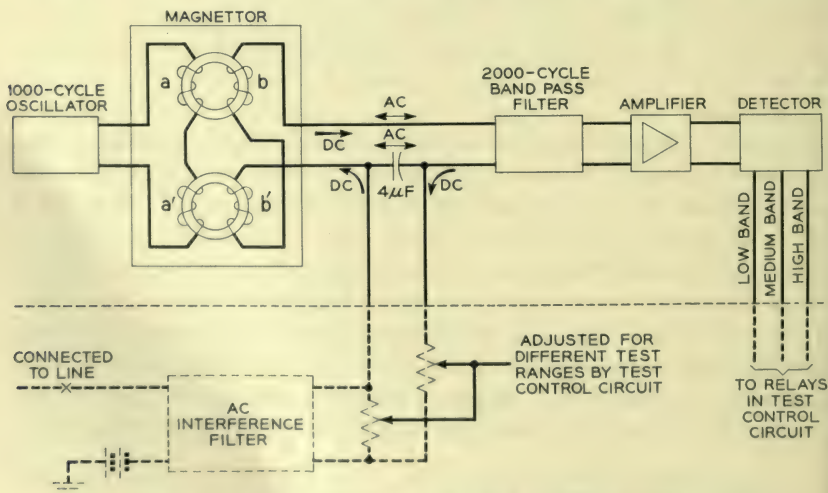


Fig. 10 — Block diagram of line insulation test circuit.

a resultant output which contains even harmonics of the input voltage. The second harmonic is selected by means of a filter for use as an indicator of the presence of leakage current.

The second harmonic is amplified by a three stage negative feedback vacuum tube amplifier, whose output is applied to the grids of three thyratrons. Adjustments are made, as described in the following, so that one of the thyratrons conducts when the insulation resistance is less than the value selected for the test, an additional thyatron conducts if the insulation resistance is one-half of this value or less, and all three conduct if the resistance is one-fourth or less of the selected value. Three relays, whose windings are connected in the anode circuits of the thyratrons, are actuated when the associated thyratrons conduct and cause a record of the trouble to be made, or cause the control circuit to select another line if the insulation resistance is above the range of interest.

In order to calibrate the detecting circuit, a test resistor of 160,000 ohms is connected to the magnetor by operation of a key. The amplifier gain is then adjusted so as to just cause conduction of the thyatron

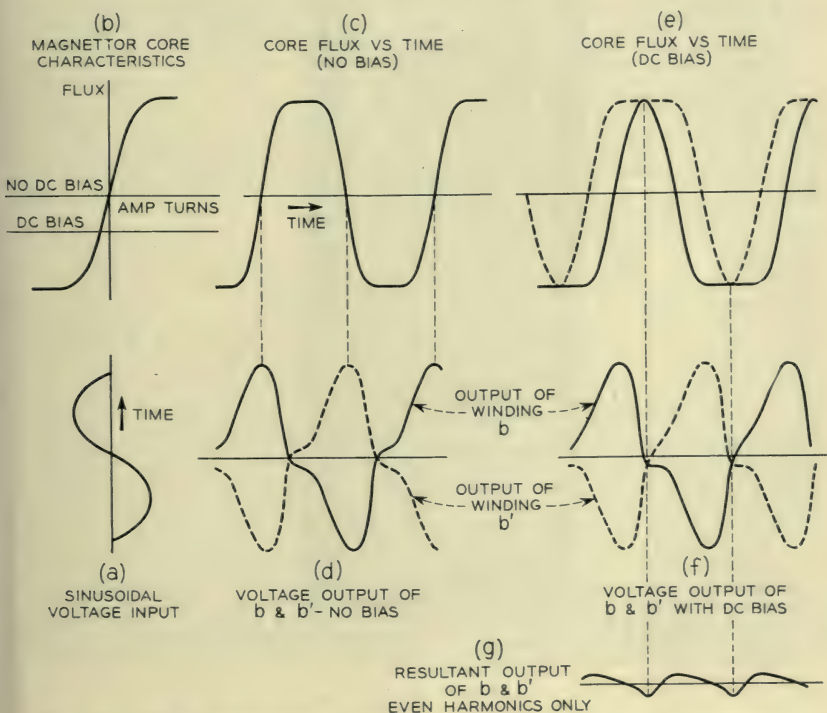


Fig. 11 — Graphical representation of voltage and flux in the magnetor.

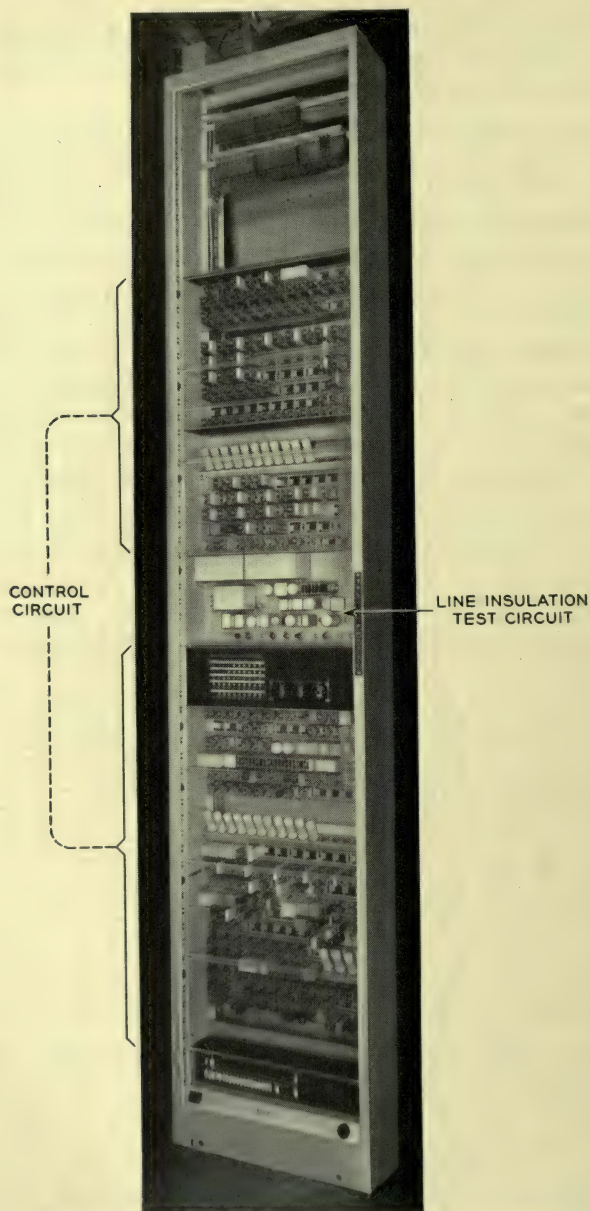


Fig. 12 — Front view of line insulation test frame.

which responds to the lowest insulation resistance value. Test resistors of 320,000 ohms and 640,000 ohms are then substituted and the grid bias of the other thyratrons adjusted so that they just conduct on the proper value of resistance. This procedure calibrates the device for one range, using a shunt and series resistor which corresponds to this range.

Facilities are provided for substituting suitable test resistors for checking the calibration of other ranges, with the shunt and series resistors for the range connected to the magnetor.

EQUIPMENT FEATURES

The apparatus components of the test control circuit and the line insulation test circuit are assembled and wired in an 11-foot bay, 23 inches wide. Fig. 12 is a front view of the test frame. Fig. 13 is an enlarged view of the line insulation test circuit equipment and the control panel which includes features for checking the accuracy of the test circuit and for calibrating it. When the teletype method of recording failures

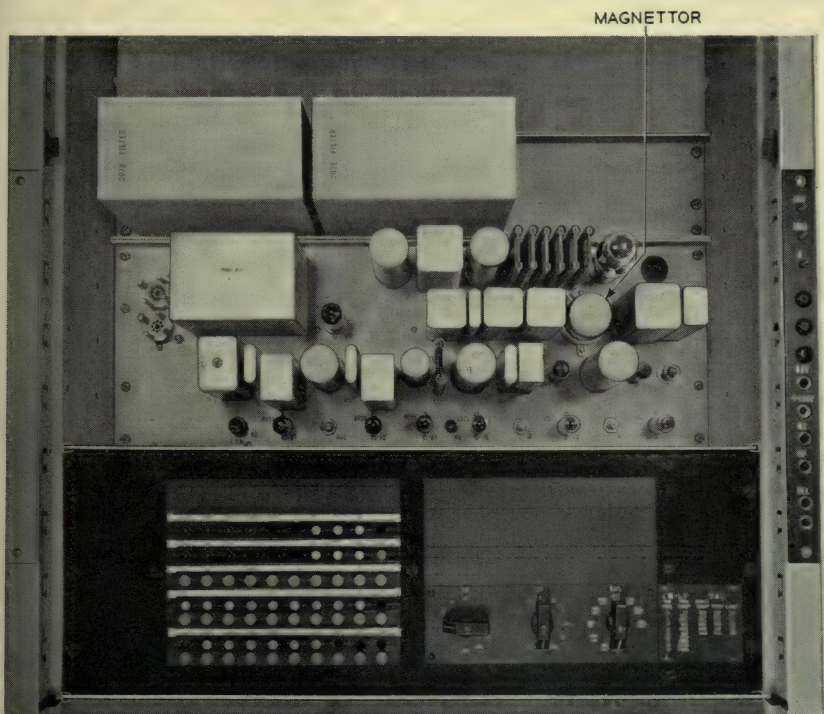


Fig. 13 — Front view of control and test panel (at bottom) and line insulation test unit (center) for No. 5 crossbar.

is used, the teletype transmitting equipment common to all line insulation test frames in a building is mounted in a separate 23-inch bay and occupies about one-third of the vertical space.

CONCLUSION

The initial installation of the automatic line insulation test equipment in a No. 5 crossbar office has been in operation for more than a year and the maintenance advantages of remote control of starting and automatic testing and recording have been fully demonstrated. It is expected that any future developments of line insulation test equipment undertaken for other switching systems will follow this same general pattern.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the assistance of R. C. Avery, F. E. Blount and D. H. MacPherson in the preparation of the technical descriptions and analyses presented in this paper.

Theory of Magnetic Effects on the Noise in a Germanium Filament

By HARRY SUHL

(Manuscript received October 10, 1952)

A magnetic field will influence the current noise in a germanium filament. This fact bears out the hypothesis that at least part of the noise arises from minority carriers emitted in random bursts and recombining at the surfaces. A quantitative theory of this effect is given.

INTRODUCTION

In a series of fundamental experiments, H. C. Montgomery¹ has established that minority carriers play an important part in the current-noise associated with semiconductors. He found that on the one hand, the noise voltage is usually proportional to the biasing current, suggesting fluctuations in the conductivity, and hence the carrier concentration. On the other hand the spectrum of the noise suggested a rather coarse-grained time variation, not likely to be caused by fluctuations in the normal carrier density. One might conclude, therefore, that the noise is caused by a distribution of sources emitting or absorbing minority carriers in random bursts. Such carriers would be subject to the same laws of motion and of recombination as intentionally injected carriers. Montgomery was, in fact, able to verify that the noise along a filament showed marked correlation over a distance roughly equal to that through which minority carriers could drift in the biasing field before recombination.

W. Shockley has pointed out another corollary of this theory: A magnetic field transverse to the filament should have a pronounced effect on the noise. This conclusion, too, Montgomery was able to verify experimentally.¹ His results are in good qualitative agreement with theory. Complete quantitative agreement was perhaps not to be expected, since technical difficulties prevented attainment of the idealized conditions assumed by the theory. This paper gives an account of that

theory. On it are based the computed curves in Montgomery's paper showing the change in noise power with magnetic field.

To see how such a change comes about, we imagine the magnetic field applied normal to one pair of the long faces of a rectangular filament. This field, and the longitudinal drift current used to measure the noise, yield a sidewise thrust on the carriers, directed at right angles to the other pair of long sides. As a result the density distribution over the cross section is distorted, the minority carriers tend to accumulate near one of those sides, while the neighborhood of the opposite side is depleted. But for the usual conditions the recombination of carriers occurs mainly near the surface, and is proportional to their density there. Hence the magnetic field will change their lifetime.^{2, 3} Clearly the amount of noise is dependent on the length of time carriers are able to contribute to the change in conductivity, that is, dependent on their lifetime. Therefore, the magnetic field should change the noise power. In simple extreme cases one can even make a semiquantitative argument for the maximum variation to be expected on the basis of such considerations.¹

FORMULATION OF THE PROBLEM

In order to make an exact calculation, we require a few preliminaries: The conductivity g is supposed to undergo a small time-dependent fluctuation $\Delta g(t)$ about its mean value.

The fluctuation arises from certain sources each of which, for macroscopic purposes, may be considered to emit a noise-current $J(t)$ of minority carriers. Thus in a small time-interval dt' near t' the excess charge injected is $J(t') dt'$. This charge decays by recombination. Let $r(t - t')$ denote the fraction of carriers left over at time $t(>t')$. Then at time t there remains a charge $r(t - t') J(t') dt'$ of the original injection. Now provided the excess density is small compared with the mean density, $\Delta g(t)$ is proportional to the excess charge at time t , due to all the previous emissions added together. Therefore

$$\Delta g(t) \propto \int_{-\infty}^t r(t - t') J(t') dt'. \quad (1)$$

In practice we do not literally plot $\Delta g(t)$ as a function of t , but rather its frequency component $\Delta g(f)$ in a narrow range df of frequencies near f . In other words, we single out for observation the contribution to Δg from that part $J(f)$ of the injected current $J(t')$ which varies as $e^{-2\pi i f t}$. Suppose now that $1/f$ is large compared with the time over which $r(t)$

is appreciably different from zero (that is, let $1/f$ be much greater than the lifetime). Then, in the integral (i), $r(t - t')$ will have gone from unity to zero long before $J(f)e^{-2\pi ift}$ has changed appreciably from its value at $t' = t$. Therefore, for purposes of observation at frequencies much smaller than the reciprocal lifetime, we can rewrite (1) as

$$\begin{aligned}\Delta g(t) &\propto J(t) \int_{-\infty}^t r(t - t') dt' \\ &= J(t) \int_0^{\infty} r(t) dt.\end{aligned}\tag{2}^*$$

The integral in (2) can be interpreted as the average lifetime of carriers. For, by definition, the rate of recombination at time t is $-dr(t)/dt$, so that $-(dr/dt)dt$ is the number of carriers recombining between time t , $t + dt$. Hence the average lifetime is

$$\begin{aligned}\tau &= -\int_0^{\infty} t \frac{dr(t)}{dt} dt = -[tr(t)]_0^{\infty} + \int_0^{\infty} r(t) dt \\ &= \int_0^{\infty} r(t) dt\end{aligned}$$

since $tr(t) \rightarrow 0$ as $t \rightarrow \infty$.

If $1/f$ is not large compared with τ one cannot simplify the integral (1) in this way. One then has to consider separately each frequency component $\Delta g(f)e^{-2\pi ift}$ due to the current $J(f)e^{-2\pi ift}$. Then

$$\begin{aligned}\Delta g(f)e^{-2\pi ift} &= J(f) \int_{-\infty}^t r(t - t')e^{-2\pi ift'} dt' \\ &= J(f)e^{-2\pi ift} \int_0^{\infty} r(t')e^{2\pi ift'} dt'\end{aligned}$$

or

$$\Delta g(f) = J(f)\tau(f)$$

where

$$\tau(f) = \int_0^{\infty} r(t')e^{2\pi ift'} dt'.$$

The calculation of $\tau(f)$ is more complicated than that of $\tau = \tau(0)$, and

* From here on the equality sign will replace the proportionality sign. The resulting change of units is of no consequence in the final results which are only concerned with ratios of conductivity modulations.

at the present time the experimental situation does not call for refinements of this kind. Therefore we shall restrict ourselves to the calculation of τ .

To evaluate τ it is not necessary to consider a time-dependent case at all. In our experiment, it is the mean square conductivity fluctuation which is actually observed. Hence from (2)

$$\langle \Delta g^2 \rangle = \langle J^2(t) \rangle \tau^2.$$

If the emission processes are stationary in time, $\langle J^2(t) \rangle$ is time independent:

$$\langle \Delta g^2 \rangle = \langle J_0^2 \rangle \tau^2.$$

Now τ can be written as

$$\int_{-\infty}^t r(t - t') dt',$$

which is simply the total concentration at the present time t due to a constant injection from time $-\infty$ to the present.

Therefore the problem is reduced to finding the total carrier concentration in the filament due to a distribution of sources of constant strength $\sqrt{\langle J_0^2 \rangle}$.

Let $w(x, y, z; x_1, y_1, z_1)$ denote the carrier concentration at x, y, z due to a steady unit source at x_1, y_1, z_1 . Then the total carrier concentration is

$$\tau(x_1, y_1, z_1) = \int w(x, y, z; x_1, y_1, z_1) dx dy dz.$$

The reason for the dependence on x_1, y_1, z_1 , is that the recombination process takes place largely on the surface. Therefore a source near the surface will yield a smaller concentration than one well inside the filament. (Volume recombination will be neglected throughout this paper.)

The mean square conductivity modulation due to many statistically independent sources at x_r, y_r, z_r ($r = 1, 2 \dots$) is then

$$\langle (\Delta g)^2 \rangle = \Sigma \tau^2(x_r, y_r, z_r) \langle J^2(x_r, y_r, z_r) \rangle.$$

The behavior of w is governed by the diffusion equation, subject to the boundary conditions expressing the recombination process, and subject to a suitable singularity at x_r, y_r, z_r , expressing the injection of a unit current. But in two and three dimensions the solution is not available in closed form, or at any rate not in terms of the elementary transcendental functions. The infinite series for the solution is not easy to

handle computationally. It is therefore desirable to simplify the experimental conditions to a point where the problem becomes almost one-dimensional. A solution for w can then be found in closed form.

Consider a very long uniform rectangular filament with one pair of sides very much wider than the other pair. Suppose that the y and z directions are respectively parallel to the wide sides and to the length of the filament (Fig. 1).

Consider sources located anywhere on a plane $x = \xi$, which is parallel to the wide sides of the filament. If the recombination properties of the filament are uniform in the $y - z$ directions, the lifetime due to a unit source anywhere in that plane is independent of the location of the source on that plane, and depends only on ξ . Hence the conductivity modulation due to sources of strength

$$J(\xi, y_r, z_r) \qquad (r = 1, 2 \dots)$$

in that plane is simply

$$\tau(\xi) \sum_r J(\xi, y_r, z_r)$$

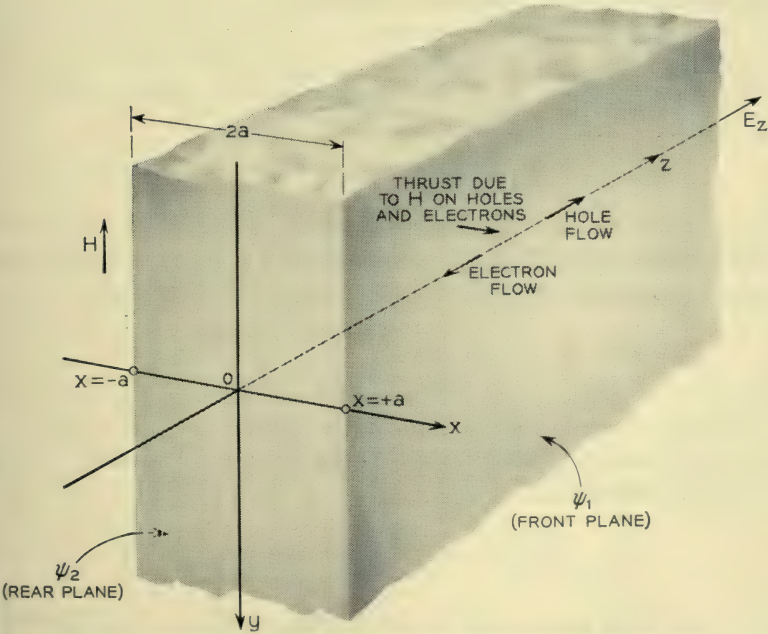


Fig. 1 — Geometry of the filament, and disposition of the fields.

and is the same as that due to an infinite source of strength

$$\sum_r J(\xi, y_r, z_r)$$

uniformly distributed over the plane x - ξ .

But the density w due to such a source will be a function of x and ξ only, and will be the solution of the one-dimensional diffusion equation. Hence for a geometry approaching that of figure 1 sufficiently closely, the problem is one-dimensional.

The Evaluation of τ

We now have to write down the one-dimensional diffusion equation in the presence of a magnetic field along 0_y , which combines with the drift velocity of the carriers so as to force them towards one of the surfaces $x = \pm a$. If F_x is the effective field arising in this manner, and D is the diffusion coefficient, the equation is

$$D \frac{d^2 w}{dx^2} - \mu F_x \frac{dw}{dx} = 0, \quad (3)$$

which expresses the fact that the diffusion current $-Dq \, dw/dx$ plus the drift current $q\mu F_x w$ must be constant since the carrier density cannot build up indefinitely. μ is the mobility of the minority carriers: μ_n for electrons, μ_p for holes. As is shown elsewhere² the effective field F_x is given by

$$F_x = (\theta_n + \theta_p) E_z = \theta E_z,$$

where E_z is the biasing field causing the drift current, θ_n , θ_p are the Hall angles for electrons and holes, respectively. If μ_n , μ_p are the electronic and hole-mobilities, and if the magnetic field is not too large,

$$\theta = \theta_n + \theta_p = 10^{-8}(\mu_n + \mu_p)H,$$

where H is in oersteds, and the mobilities are in $\text{cm}^2/\text{volt-second}$, and θ is in radians. (Strictly speaking, the diffusion current is not in the direction of the density gradient when a magnetic field is present.⁴ As the result mixed derivatives

$$\frac{\partial^2 w}{\partial x \partial y}$$

occur in the diffusion equation. But in the reduction to one dimension these terms integrate out. All that remains is a small correction to D ,

negligible for ordinary values of H .) It is convenient to specify a dimensionless parameter in the same notation as H. C. Montgomery.

$$\Phi = \frac{2a\mu F_x}{D} = \frac{2a\theta E_z}{D/\mu}$$

By the Einstein Relation $D/\mu = kT/q$ this may be written

$$= \frac{2a\theta E_z}{kT/q}$$

where q is the absolute value of the electronic charge. Φ is the ratio of the voltage corresponding to the transverse field to the thermal voltage kT/q . In terms of Φ , equation (3) can be rewritten

$$\frac{d^2w}{dx^2} - \frac{\Phi}{2a} \frac{dw}{dx} = 0 \quad (4)$$

The integral of this equation has the form

$$w = Ae^{\Phi(x/2a)} + B \quad (5)$$

where A and B are two constants. Because of the existence of a singularity at $x = x_0$, say, the constants A , B take on different values for $x < x_0$ and $x > x_0$. To see what these values are, we first write the solution (5) in the form

$$w_1 = A_1 e^{\Phi(x-x_0)/2a} + B_1 \quad x > x_0,$$

$$w_2 = A_2 e^{\Phi(x-x_0)/2a} + B_2 \quad x < x_0.$$

At $x = x_0$ the density w must be continuous. Hence

$$A_1 + B_1 = A_2 + B_2 \quad (6)$$

Further, the discontinuity at x_0 must be such that the difference of the currents on the two sides is just unity, the strength of the injected current. Now the total current is

$$-D \left(\frac{dw}{dx} - \frac{\Phi}{2a} w \right)$$

(i.e., the diffusion current plus the drift current), and w is continuous at x_0 . Hence the difference between the current on the two sides of x_0 is

$$\begin{aligned} 1 &= -D \lim_{h \rightarrow 0} \left(\left(\frac{dw_1}{dx} \right)_{x_0+h} - \left(\frac{dw_2}{dx} \right)_{x_0-h} \right) \\ &= \frac{\Phi}{2a} D(A_2 - A_1). \end{aligned} \quad (7)$$

So far we have two relations (6) and (7) between the four constants A_1, A_2, B_1, B_2 . To determine these constants, we need two more equations. These are supplied by the boundary conditions. We assume that the recombination rate is proportional to the density at the boundaries ($x = \pm a$). The recombination then has the formal appearance of a current through the surface, the factor of proportionality s playing the part of a "recombination velocity."^{2, 3} That current through the surface must equal the current arriving at the surface from the interior of the filament, under steady-state conditions. The boundary conditions are thus:

$$\begin{aligned} -D \left(\frac{dw_1}{dx} - \frac{\Phi}{2a} w_1 \right)_{x=+a} &= s_1 w(+a), \\ -D \left(\frac{dw_2}{dx} - \frac{\Phi}{2a} w_2 \right)_{x=-a} &= -s_2 w(-a), \end{aligned}$$

where s_1, s_2 are the surface recombination velocities of the faces $x = +a, -a$ respectively. The minus sign on the right-hand side of the second of these equations is due to the fact that the recombination current at $x = -a$ is along the $-x$ direction. Defining the recombination parameters

$$\psi_{1, 2} = \frac{s_{1, 2} a}{D}$$

and substituting the solutions w_1, w_2 in the last two equations we obtain

$$\begin{aligned} B_1 \frac{\Phi}{2} &= \psi_1 (A_1 e^{\Phi(1-x_0/a)/2} + B_1) \\ B_2 \frac{\Phi}{2} &= \psi_2 (A_2 e^{-\Phi(1+x_0/a)/2} + B_2) \end{aligned} \quad (8)$$

Equations (6), (7) and (8) suffice to determine A_1, A_2, B_1, B_2 . We easily find that

$$\begin{aligned} A_1 &= -e^{-\Phi(1-x_0/a)/2} \frac{2a}{\Phi D} \frac{\alpha_1(1 - \alpha_2 e^{\Phi(1+x_0/a)/2})}{\Delta} \\ A_2 &= +e^{+\Phi(1+x_0/a)/2} \frac{2a}{\Phi D} \frac{\alpha_2(1 + \alpha_1 e^{-\Phi(1-x_0/a)/2})}{\Delta} \\ B_1 &= -\frac{2a}{\Phi D} \frac{(1 - \alpha_2 e^{\Phi(1+x_0/a)/2})}{\Delta} \\ B_2 &= -\frac{2a}{\Phi D} \frac{(1 + \alpha_1 e^{-\Phi(1-x_0/a)/2})}{\Delta} \end{aligned} \quad (9)$$

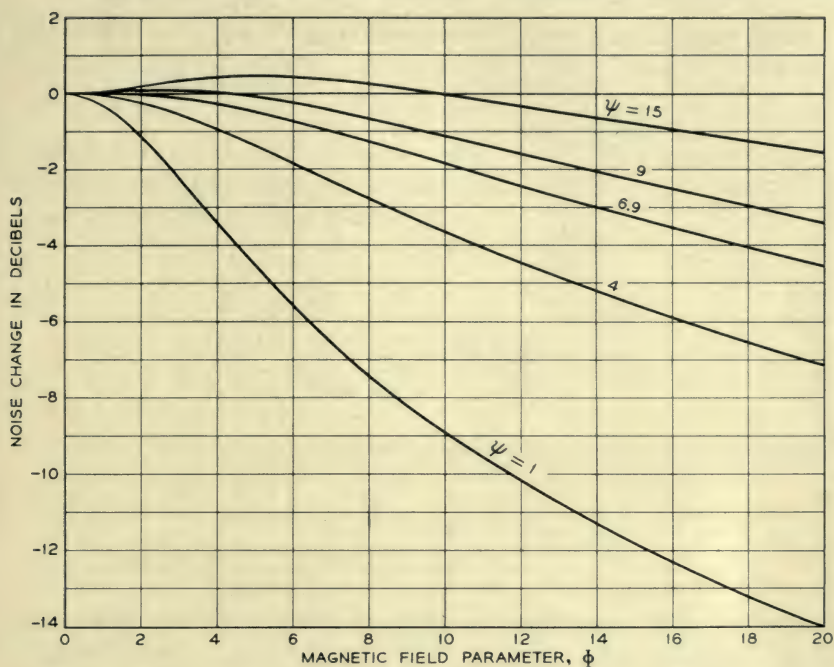


Fig. 2 — Noise change for a filament with equal surfaces as a function of Φ for various ψ . Surface generation.

where

$$\alpha_1 = \left(\frac{\Phi}{2} - \psi_1 \right) / \psi_1 ; \quad \alpha_2 = \left(\frac{\Phi}{2} + \psi_2 \right) / \psi_2 ;$$

$$\Delta = \alpha_1 e^{-\Phi(1-x_0/a)/2} + \alpha_2 e^{\Phi(1+x_0/a)/2}.$$

Now we are in a position to calculate $\tau_H(x_0)$, the life-time in a magnetic field H (contained in Φ), due to an emission at x_0 :

$$\begin{aligned} \tau_H(x_0) &= \int_{-a}^{x_0} w_2(x, x_0) dx + \int_{x_0}^{+a} w_1(x, x_0) dx \\ &= (B_1 + B_2)a + (B_2 - B_1)x_0 + \frac{2a}{\Phi} (A_2 - A_1) \end{aligned} \quad 10$$

$$\frac{2a}{\Phi} [A_1 e^{\Phi(1-x_0/a)/2} - A_2 e^{-\Phi(1+x_0/a)/2}].$$

Special Cases

With the help of (9) and (10) we can now examine special cases.

1. Surface Emission

When only the surface $x = +a$ is emitting, we set $x_0 = +a$ in (10) and get

$$\begin{aligned}\tau_H(+a) &= 2B_2\tau + \frac{2a}{\Phi} A_2(1 - e^{-\Phi}) \\ &= \frac{4a^2}{\Phi D} (1 + \alpha_1) \frac{\left[1 - \alpha_2 \frac{e^{\Phi}}{\Phi} (1 - e^{-\Phi})\right]}{\alpha_1 + \alpha_2 e^{\Phi}} \\ &= -\frac{2a^2}{D\psi_1} \frac{\left(1 - \alpha_2 \frac{(e^{\Phi} - 1)}{\Phi}\right)}{\alpha_1 + \alpha_2 e^{\Phi}}.\end{aligned}$$

Similarly, when only $-a$ is emitting, we get

$$\tau_H(-a) = -\frac{2a^2}{D\psi_2} \frac{\alpha_1 \frac{e^{-\Phi} - 1}{\Phi} - 1}{\alpha_1 e^{-\Phi} + \alpha_2}.$$

But in an actual experiment, both faces will be emitting, with mean square strengths $\langle J_1^2 \rangle$, $\langle J_2^2 \rangle$ say. The quantity that is then measured

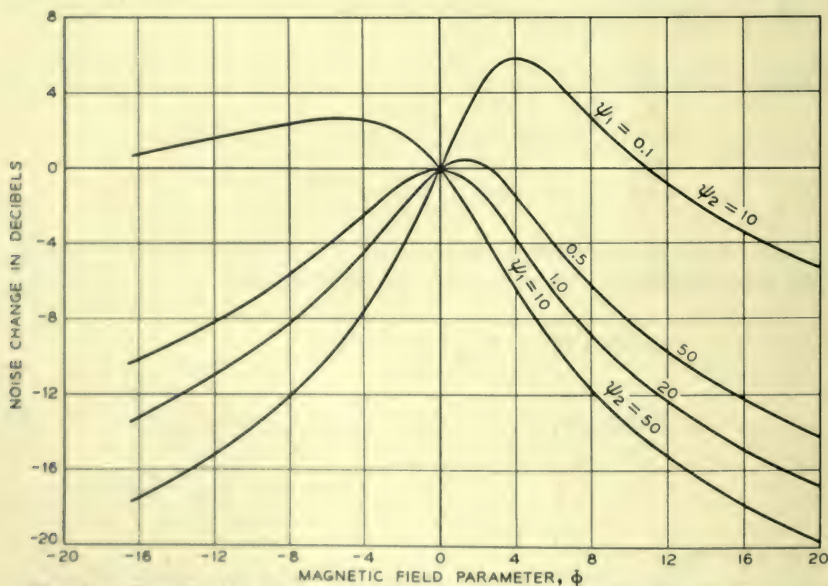


Fig. 3 — Contribution to the noise change from a unit source at the plane $x = +a$.

is the ratio

$$N_H = \frac{\Delta g_H^2}{\Delta g_{H=0}^2} = \frac{\langle J_1^2 \rangle \tau_H^2(+a) + \langle J_2^2 \rangle \tau_H^2(-a)}{\langle J_1^2 \rangle \tau_{H=0}^2(+a) + \langle J_2^2 \rangle \tau_{H=0}^2(-a)}. \quad (11)$$

Therefore we also need the lifetimes at zero H (that is, zero Φ). But at $\Phi = 0$, the τ 's are indeterminate, and we therefore have to take limits. Expansion in powers of Φ shows that

$$\lim_{\Phi \rightarrow 0} \tau_H(+a) = \frac{a^2}{D\psi_1} \frac{\left(1 + \frac{1}{\psi_2}\right)}{1 + \frac{1}{2}\left(\frac{1}{\psi_1} + \frac{1}{\psi_2}\right)},$$

$$\lim_{\Phi \rightarrow 0} \tau_H(-a) = \frac{a^2}{D\psi_2} \frac{\left(1 + \frac{1}{\psi_1}\right)}{1 + \frac{1}{2}\left(\frac{1}{\psi_1} + \frac{1}{\psi_2}\right)}.$$

Thus we finally get

$$N_H = \frac{\Delta g_H^2}{\Delta g_0^2} = 4 \left(1 + \frac{1}{2}\left(\frac{1}{\psi_1} + \frac{1}{\psi_2}\right)\right)^2$$

$$\frac{\langle J_1^2 \rangle \left[\frac{1 - \alpha_2 \left(\frac{e^\Phi - 1}{\Phi} \right)^2}{\alpha_1 + \alpha_2 e^\Phi} \right]^2}{\psi_1^2} + \frac{\langle J_2^2 \rangle \left[\frac{\alpha_1 \left(\frac{1 - e^{-\Phi}}{\Phi} \right) + 1}{\alpha_1 e^{-\Phi} + \alpha_2} \right]^2}{\psi_2^2} \quad (12)$$

$$\frac{\langle J_1^2 \rangle \left(1 + \frac{1}{\psi_2}\right)^2}{\psi_1^2} + \frac{\langle J_2^2 \rangle \left(1 + \frac{1}{\psi_1}\right)^2}{\psi_2^2}$$

There remains one small difficulty in the way of comparing experiment with theory: We do not know $\langle J_1^2 \rangle$, $\langle J_2^2 \rangle$. As suggested by H. C. Montgomery, we are able to overcome this difficulty as follows: We first draw a number of curves of

$$\frac{\tau_H^2(+a)}{\tau_0^2(+a)}, \quad \frac{\tau_H^2(-a)}{\tau_0^2(-a)} \text{ versus } \Phi,$$

for various sets of parameters (ψ_1, ψ_2) . (See figures 3, 4). Then we contrive to match a linear superposition

$$c_1 \frac{\tau_H^2(+a)}{\tau_0^2(+a)} + c_2 \frac{\tau_H^2(-a)}{\tau_0^2(-a)}$$

where $c_1 + c_2 = 1$ to the experimental curve. This will be possible only for one particular set of values (ψ_1, ψ_2) . From c_1 and $c_2 = 1 - c_1$ and

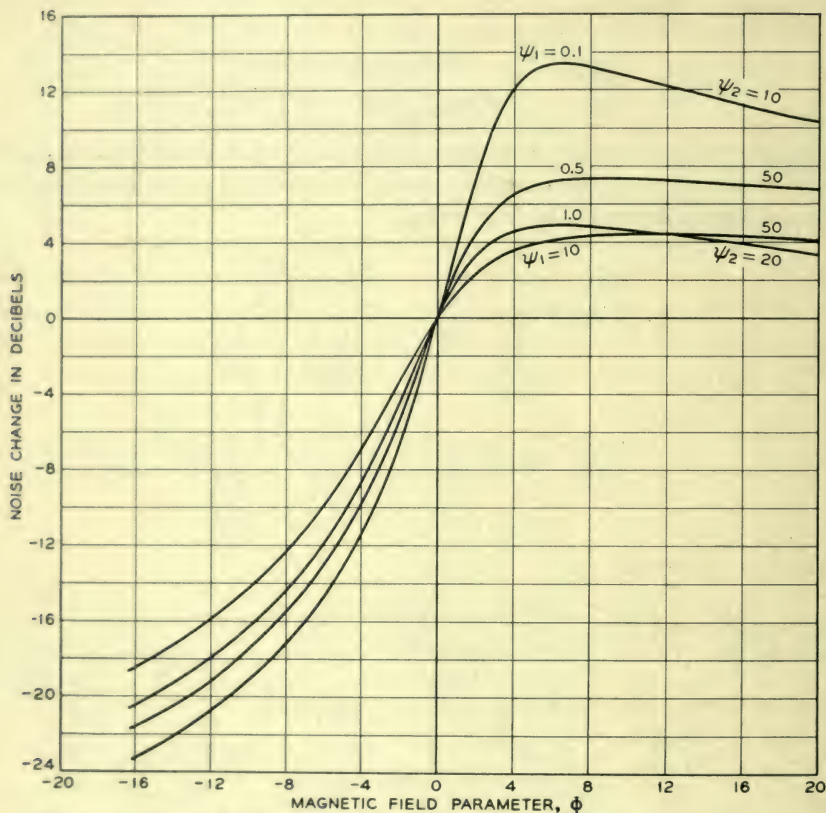


Fig. 4 — Contribution to the noise change from a unit source at $x = -a$.

from (ψ_1, ψ_2) we can determine the ratio of $\langle J_1^2 \rangle \langle J_2^2 \rangle$, for, having matched the experimental curve, we know that

$$c_1 \frac{\tau_H^2(+a)}{\tau_0^2(+a)} + (1 - c_1) \frac{\tau_H^2(-a)}{\tau_0^2(-a)} = \frac{\langle J_1^2 \rangle \tau_H^2(+a) + \langle J_2^2 \rangle \tau_H^2(-a)}{\langle J_1^2 \rangle \tau_0^2(+a) + \langle J_2^2 \rangle \tau_0^2(-a)},$$

which is satisfied for all Φ if

$$\frac{c_1}{\tau_0^2(+a)} = \frac{\langle J_1^2 \rangle}{\langle J_1^2 \rangle \tau_0^2(+a) + \langle J_2^2 \rangle \tau_0^2(-a)},$$

$$\frac{(1 - c_1)}{\tau_0^2(-a)} = \frac{\langle J_2^2 \rangle}{\langle J_1^2 \rangle \tau_0^2(+a) + \langle J_2^2 \rangle \tau_0^2(-a)}.$$

These equations are self-consistent and give

$$\frac{c_1}{1 - c_1} = \frac{\tau_0^2(+a)\langle J_1^2 \rangle}{\tau_0^2(-a)\langle J_2^2 \rangle} = \frac{\langle J_1^2 \rangle}{\langle J_2^2 \rangle} \frac{\left(1 + \frac{1}{\psi_2}\right)^2 \psi_2^2}{\left(1 + \frac{1}{\psi_1}\right)^2 \psi_1^2}.$$

Thus the match of theory to experiment actually provides information about the relative emissivities of the surfaces.

When the surfaces are equal, $\psi_1 = \psi_2 = \psi$; $\langle J_1^2 \rangle = \langle J_2^2 \rangle$ and the result 12 simplifies to

$$N_H = \frac{\left(1 + \frac{1}{\psi}\right)^2 + \left(\frac{2}{\Phi} - \coth \frac{\Phi}{2}\right)^2}{\left(1 + \frac{\Phi}{2\psi} \coth \frac{\Phi}{2}\right)^2}. \quad (13)$$

$10 \log N_H$ for this case is plotted in figure 2.

It is interesting to see how N_H for equal surfaces varies for small values of Φ . $2/\Phi - \coth \Phi/2$ is regular at $\Phi = 0$ and varies as $-\Phi/6$ there. Hence the initial variation is as

$$\begin{aligned} N_{H\text{small}} &= 1 + \frac{1}{4\left(1 + \frac{1}{\psi}\right)^2} \left(\frac{\Phi^2}{9} - \frac{2}{3}\frac{\Phi^2}{\psi} \left(1 + \frac{1}{\psi}\right)\right) \\ &= 1 + \frac{\Phi^2}{12\left(1 + \frac{1}{\psi}\right)^2} \left(\frac{1}{3} - \frac{2}{\psi} \left(1 + \frac{1}{\psi}\right)\right). \end{aligned}$$

Hence the curve rises or falls initially according as

$$\frac{1}{3} \gtrless 2\frac{1}{\psi} \left(1 + \frac{1}{\psi}\right).$$

The noise therefore increases initially if

$$\sqrt{\frac{5}{12}} - \frac{1}{2} > \frac{1}{\psi} \quad \text{or} \quad \psi > 6.9 \text{ approximately}$$

and falls initially if

$$\psi < 6.9.$$

2. Volume Generation.

At first sight it may seem that if volume generation is considered, so should volume recombination (detailed balancing). This is not necessarily so, since we are not dealing with an equilibrium situation here.

Therefore the possibility of volume generation and surface recombination cannot be discarded.

This case is somewhat more difficult. Assuming all sources to be uncorrelated and uniformly distributed throughout the interval $(-a, +a)$, we have to square expression (10) and integrate it from $x_0 = -a$ to $x_0 = +a$ in order to find $\langle \Delta g_H^2 \rangle$. (We suppose that all the sources are of equal strength $\langle J^2 \rangle$). Substituting the values of the A and B from 9 and 10, we get, after some obvious cancellations

$$\tau_H(x_0) = \frac{2a}{\Phi D} [x_0 + S e^{-\Phi(x_0/2a)} + T]$$

where

$$S = 2a \frac{\left(1 + \frac{1}{2} \left(\frac{1}{\psi_1} + \frac{1}{\psi_2}\right)\right)}{\alpha_1 e^{-\Phi/2} + \alpha_2 e^{\Phi/2}},$$

$$T = \frac{2a}{\Phi} \left[\frac{\Phi}{2} \frac{\alpha_1 e^{-\Phi/2} - \alpha_2 e^{+\Phi/2}}{\alpha_1 e^{-\Phi/2} + \alpha_2 e^{+\Phi/2}} - 2 \frac{\alpha_1 \alpha_2 \sinh \frac{\Phi}{2}}{\alpha_1 e^{-\Phi/2} + \alpha_2 e^{\Phi/2}} - 1 \right].$$

Hence

$$\begin{aligned} \langle \Delta g_H^2 \rangle &= \langle J \rangle \int_{-a}^{+a} \tau_H^2(x_0) dx_0 \\ &= \frac{4a^2}{\Phi^2 D^2} \left[\frac{2}{3} a^3 + 2T^2 a + \frac{8TSa}{\Phi} \sinh \frac{\Phi}{2} \right. \\ &\quad \left. + \frac{2aS^2}{\Phi} \sinh \Phi + \frac{8Sa^2}{\Phi} \left(\frac{\sinh \frac{\Phi}{2}}{\frac{\Phi}{2}} - \cosh \frac{\Phi}{2} \right) \right]. \end{aligned} \quad (14)$$

Before proceeding with this general case, we first consider the limiting case $\psi_1 = \psi_2 = \infty$, when $\alpha_1 = -1$, $\alpha_2 = +1$. Then

$$T = -a \coth \frac{\Phi}{2} \quad S = \frac{a}{\sinh \frac{\Phi}{2}}$$

and

$$\int_{-a}^{+a} \tau_H^2(x_0) = \frac{8a^5}{\Phi^2 D^2} \left[\frac{1}{3} + \coth^2 \frac{\Phi}{2} - 6 \frac{\coth \frac{\Phi}{2}}{\Phi} + \frac{8}{\Phi^2} \right].$$

To find $N_H = \langle \Delta g_H^2 \rangle / \langle \Delta g_{H=0}^2 \rangle$ we need the limit of $\int_{-a}^{+a} \tau_H^2$ as $\Phi \rightarrow 0$. After some tedious algebra, we find this limit to be

$$\tau_{H=0}^2 = \frac{4a^5}{15D^2}$$

so that

$$(N_H)_{\psi_1, \psi_2 \rightarrow \infty} = \frac{\int \tau_H^2}{\int \tau_0^2} = \frac{30}{\Phi^2} \left[\frac{1}{3} + \coth^2 \frac{\Phi}{2} - \frac{6 \coth \frac{\Phi}{2}}{\Phi} + \frac{8}{\Phi^2} \right]. \quad (15)$$

In the general case we can again take the limit of (14) as $\Phi \rightarrow 0$ in order to determine N_H , but this would be too tedious. Instead, we solve the diffusion equation directly when $\Phi = 0$. The equation is then simply

$$\frac{d^2 w}{dx^2} = 0$$

and the solution subject to the correct boundary conditions and allowing for a steady unit injection at x_0 is

$$\begin{aligned} w_1 &= A_1(x - x_0) + B & x > x_0 \\ w_2 &= A_2(x - x_0) + B & x < x_0. \end{aligned}$$

where

$$\begin{aligned} A_1 &= -\frac{1}{D} \psi_1 \left[1 + \psi_2 \left(1 + \frac{x_0}{a} \right) \right] / (\psi_1 + \psi_2 + 2\psi_1\psi_2), \\ A_2 &= \frac{1}{D} \psi_2 \left[1 + \psi_1 \left(1 - \frac{x_0}{a} \right) \right] / (\psi_1 + \psi_2 + 2\psi_1\psi_2), \\ B &= \frac{a}{D} \left[1 + \psi_1 \left(1 - \frac{x_0}{a} \right) \right] \left[1 + \psi_2 \left(1 + \frac{x_0}{a} \right) \right] / (\psi_1 + \psi_2 + 2\psi_1\psi_2). \end{aligned}$$

We now have

$$\begin{aligned} \tau_{H=0}(x_0) &= \int_{-a}^{+a} w \, dx = \int_{-a}^{x_0} w_2 \, dx + \int_{x_0}^a w_1 \, dx \\ &= -\frac{1}{2} \left[2ax_0(A_1 + A_2) + \frac{a^2 + x_0^2}{D} \right] + 2aB. \end{aligned}$$

From this we can compute

$$\int_{-a}^{+a} \tau_{H=0}^2(x_0) \, dx_0,$$

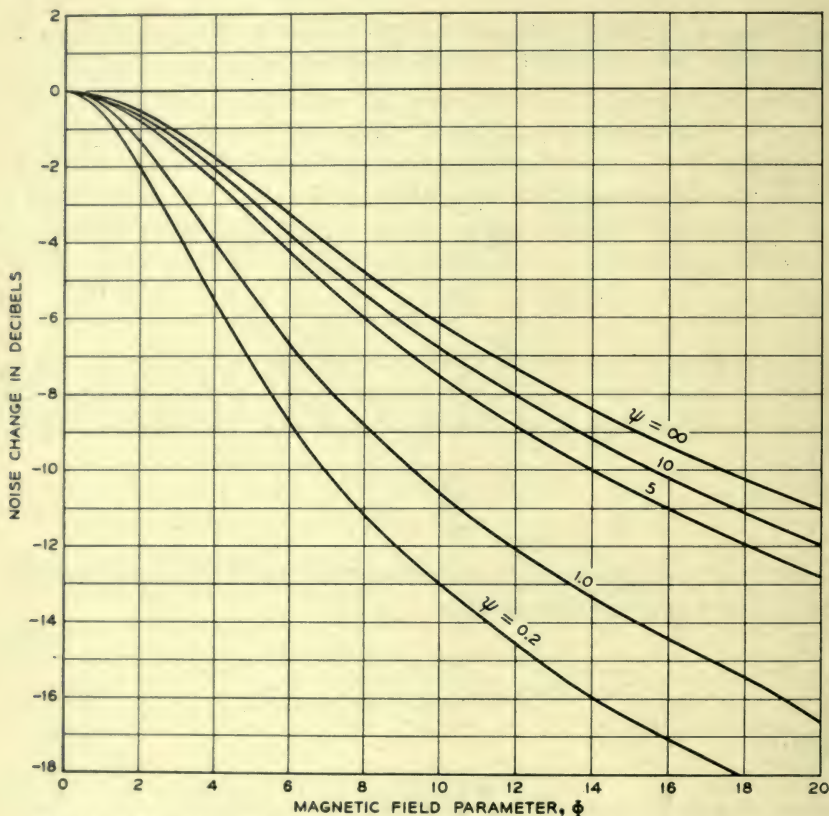


Fig. 5 — Noise change in a filament with uniform volume generation and equal surface recombination rates.

but the result is still rather complicated unless $\psi_1 = \psi_2 = \psi$. If we therefore restrict ourselves to that case we find

$$A_1 + A_2 = -\frac{\psi}{D(1 + \psi)} \cdot \frac{x_0}{a},$$

$$B = \frac{1}{2D} \left[\frac{1 + \psi}{\psi} - \frac{\psi}{1 + \psi} \frac{x_0^2}{a^2} \right]$$

and

$$D\tau_{H=0}(x_0) = a^2 \left(\frac{1}{2} + \frac{1}{\psi} \right) - \frac{x_0^2}{2}$$

and

$$\int_{-a}^{+a} \tau_{H=0}^2(x_0) dx_0 = \frac{2a^5}{D^2} \left(\frac{1}{45} + \left(\frac{1}{\psi} + \frac{1}{3} \right)^2 \right).$$

(Note that as $\psi \rightarrow \infty$, this quantity tends to $4a^5/15D^2$, as before in our limiting process.)

When $\psi_1 = \psi_2 = \psi$, S and T also simplify somewhat, and the result is

$$N_H = \frac{\langle \Delta g_H^2 \rangle}{\langle \Delta g_{H=0}^2 \rangle} = 2 \frac{\frac{2}{3} + 2T^2 + \frac{8S}{\Phi} \left(\frac{2}{\Phi} - T \right) \sinh \frac{\Phi}{2} - \frac{8S}{\Phi} \cosh \frac{\Phi}{2} + \frac{2S^2}{\Phi} \sinh \Phi}{\Phi^2 \left[\left(\frac{1}{\psi} + \frac{1}{3} \right)^2 + \frac{1}{45} \right]}$$

where now

$$S = \frac{1 + \frac{1}{\psi}}{\left(\frac{\Phi}{2\psi} \cosh \frac{\Phi}{2} + \sinh \frac{\Phi}{2} \right)},$$

$$T = \left(1 + \frac{1}{\psi} \right) \frac{\frac{\Phi}{2\psi} + \coth \frac{\Phi}{2}}{1 + \frac{\Phi}{2\psi} \coth \frac{\Phi}{2}}.$$

An alternative form is

$$N_H = \frac{4}{\frac{1}{45} + \left(\frac{1}{3} + \frac{1}{\psi} \right)^2} \left[\frac{1}{3\Phi^2} + \frac{4}{\Phi^3\Delta} \left(1 + \frac{1}{\psi} \right) \left(\frac{2}{\Phi} - \coth \frac{\Phi}{2} \right) + \frac{\left(1 + \frac{1}{\psi} \right)^2}{\Phi^2\Delta^2} \left(\left(\frac{\Phi}{2\psi} + \coth \frac{\Phi}{2} \right)^2 + \frac{2}{\psi} - \frac{2 \coth \frac{\Phi}{2}}{\Phi} \right) \right] \quad (16)$$

where

$$\Delta = 1 + \frac{\Phi}{2\psi} \coth \frac{\Phi}{2},$$

a result which correctly tends to (15) as $\psi \rightarrow \infty$. $10 \log N_H$ for various ψ is shown in Fig. 5.

ACKNOWLEDGMENTS:

The author gratefully acknowledges the help given by Dr. Shockley, who suggested the conceptual picture underlying this theory and who

independently calculated some special cases. Thanks are also due to H. C. Montgomery for many valuable discussions and to Mrs. C. A. Lambert for computing the curves.

REFERENCES

1. Montgomery, H. C., Electrical Noise in semiconductors. Bell System Tech. J., **31**, pp. 950-975, Sept., 1952.
2. Shockley, W., Electrons and Holes in Semiconductors, Chapter 3, Section 2, Van Nostrand.
3. Suhl, H., and W. Shockley, Phys. Rev., **75**, pp. 1617-1618, 1949.
4. Shockley, W., Electrons and Holes in Semiconductors, p. 299, Van Nostrand.

DC Field Distribution in a "Swept Intrinsic" Semiconductor Configuration

By R. C. PRIM

(Manuscript received January 15, 1953)

This paper contains an analysis of the dc field intensity distribution in an idealized one-dimensional n-intrinsic-p semi-conductor configuration biased in reverse. It gives some quantitative insight into the progressive penetration of the field into the intrinsic region as the magnitude of the bias voltage is increased.

INTRODUCTION

Possible applications have been suggested for semi-conductor configurations involving intrinsic regions adjacent simultaneously to n - and p -type extrinsic regions. The basic idea behind some of these proposals is that a suitably large reverse bias voltage (n -regions positive with respect to p -regions) will set up a substantial electric field in the interior of the intrinsic region. This field would sweep most of the mobile carriers out of the intrinsic material, producing a region of material ("swept intrinsic") supporting a large field and having a high resistivity.

This paper contains a dc analysis of an idealized one-dimensional swept intrinsic structure with abrupt transitions from strongly n -type to highly intrinsic to strongly p -type material. It gives some quantitative insight into the penetration of the electric field into the intrinsic region as the bias voltage is progressively increased.

FORMULATION OF PROBLEM

A one-dimensional structure will be considered having the distribution of excess of donor concentration over acceptor concentration ($N_d - N_a$) shown in Fig. 1. It will be supposed that N/n_i and P/n_i are $\gg 1$ and that a reverse bias voltage (Fig. 2) is applied between the bodies of the n - and p -type regions. (n_i denotes the thermal equilibrium con-

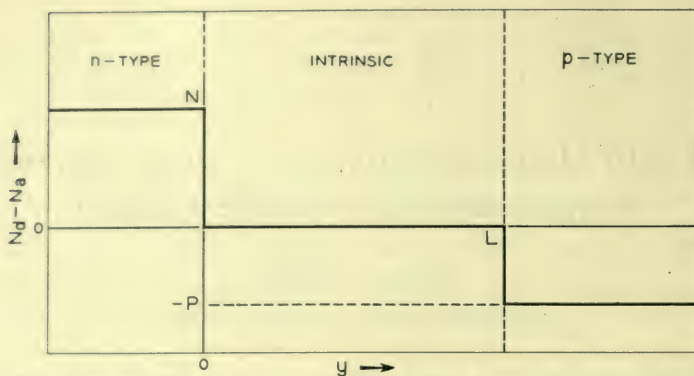


Fig. 1 — Assumed distribution of excess of donor concentration over acceptor concentration.

centration of mobile electrons — or of holes — in intrinsic material at the given temperature.)

The set of equations which (together with boundary conditions to be specified later) determines the electric field in the intrinsic region $0 < y < L$ is:

$$\frac{dE}{dy} = \frac{q}{\kappa} (p - n), \quad (1)$$

$$\frac{dp}{dy} = \frac{q}{kT} pE - \frac{i_p}{\mu_p kT}, \quad (2)$$

$$\frac{dn}{dy} = -\frac{q}{kT} nE + \frac{i_n}{\mu_n kT}, \quad (3)$$

$$\frac{d\psi}{dy} = -E, \quad (4)$$

$$\frac{di_p}{dy} = q(g - r), \quad (5)$$

$$\frac{di_n}{dy} = -q(g - r). \quad (6)$$

where E : electric field intensity, volts/m.

q : electronic charge magnitude, coulombs.

κ : absolute dielectric constant, farads/m.

p : hole concentration, m^{-3} .

n : electron concentration, m^{-3} .

k : Boltzmann's constant, joules/°K.

T : temperature, °K.

ψ : electric potential, volts.

i_p : hole current density, amps/m².

i_n : electron current density, amps/m².

μ_p : hole mobility constant, m²/volt-sec.

μ_n : electron mobility constant, m²/volt-sec.

g : rate of generation of hole-electron pairs, m⁻³ sec⁻¹.

r : rate of recombination of hole-electron pairs, m⁻³ sec⁻¹.

An order-of-magnitude comparison of the terms in (2) or (3) reveals that the currents probably have little influence on the field distribution. For example for

$$i_n \sim 5 \text{ amps/m}^2,$$

$$\mu_n kT = 1.44 \times 10^{-21} \text{ amp-m}^2,$$

$$N = 2 \times 10^{22} \text{ m}^{-3}, \text{ and } L = 10^{-3} \text{ m.}$$

$$\int_0^L \frac{i_n}{\mu_n kT} dy \sim 3 \cdot 10^{21} L \sim 3 \cdot 10^{18} \text{ m}^{-3}$$

while

$$\left| \int_0^L \frac{dn}{dy} dy \right| = n(0) - n(L) \sim N = 2 \cdot 10^{22} \text{ m}^{-3}.$$

On this basis, the current terms in (2) and (3) can be omitted without serious error. No use then has to be made of (5) and (6), so the governing equations for the intrinsic region become:

$$\frac{dE}{dy} = \frac{q}{\kappa} (p - n), \quad (1')$$

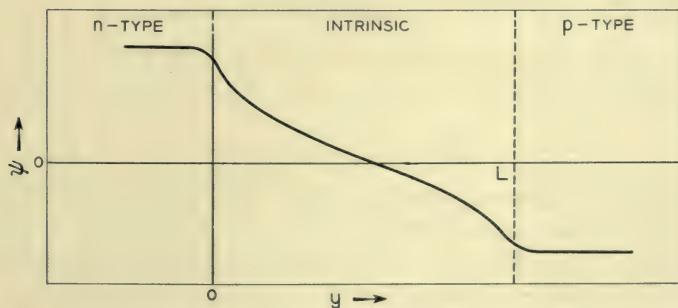


Fig. 2 — Qualitative picture of potential distribution in reverse-biased *n*-intrinsic-*p* structure.

$$\frac{dp}{dy} = \frac{q}{kT} pE, \quad (2')$$

$$\frac{dn}{dy} = -\frac{q}{kT} nE, \quad (3')$$

$$\frac{d\psi}{dy} = -E. \quad (4')$$

Equations (2')-(4') will also be used within the n -type extrinsic region ($y < 0$) and the p -type extrinsic region ($y > L$). However, for the n -type region (1') is replaced by

$$\frac{dE}{dy} = \frac{q}{\kappa} (N + p - n), \quad (1'a)$$

where N denotes the excess concentration of ionized donor over acceptor centers for $y < 0$. Similarly, for the p -type region (1') is replaced by

$$\frac{dE}{dy} = \frac{q}{\kappa} (-P + p - n), \quad (1'b)$$

where P denotes the excess concentration of ionized acceptor over donor centers for $y > L$.

In order to solve the equation set (1')-(4') for the intrinsic region $0 < y < L$, it will be necessary partially to solve the sets (1'a) (2')-(4') and (1'b) (2')-(4') governing the two extrinsic regions because only in this way can a sufficient number of appropriate boundary conditions be imposed.

Deep inside the extrinsic regions the electric field intensity will be negligible and the mobile carrier concentrations will have their equilibrium values. This leads to the conditions

$$E = 0, \quad np = n_i^2, \quad n - p = N \text{ at } y = -\infty \quad (7)$$

and

$$E = 0, \quad np = n_i^2, \quad p - n = P \text{ at } y = +\infty. \quad (8)$$

It will further be supposed that there is no infinite charge concentration at the extrinsic-intrinsic interfaces so that the electric field intensity is continuous at the interfaces. The concentration of the (local) majority carrier will also be assumed continuous at an interface. In short,

$$E \text{ and } n \text{ are continuous at } y = 0 \quad (9)$$

and

$$E \text{ and } p \text{ are continuous at } y = L. \quad (10)$$

Finally, we choose the reference level for the electric potential in the intrinsic region so that

$$\psi = 0 \text{ for } n = p \quad (11)$$

and regard the potential at the interface $y = 0$ as a prescribed parameter,

$$\psi = \frac{kT}{q} \cdot U \text{ at } y = 0. \quad (12)$$

The two conditions (11) and (12) apply directly to the solutions for the intrinsic region. The conditions (7)–(10) indirectly imply the two additional restraints necessary to determine a unique solution of (1')–(4') in $0 < y < L$.

NORMALIZED VARIABLES AND EQUATIONS

It is convenient to introduce dimensionless normalized variables before proceeding further with the mathematical analysis. As reference voltage it is natural to adopt the *Boltzmann voltage*

$$\psi_B \equiv \frac{kT}{q}, \quad (13)$$

the voltage equivalent of the mean kinetic energy of an electron at temperature T . (At room temperature the Boltzmann voltage is about 1/40 of a volt.) As reference quantity for carrier concentrations we choose the geometric mean of the majority carrier excess concentrations for the two extrinsic regions. i.e.,

$$\text{reference concentration} = (NP)^{1/2}. \quad (14)$$

The reference voltage and carrier concentration having been so chosen, it is natural to select as reference length the *mean Debye length*

$$\mathcal{L} \equiv \left[\frac{kT/q}{2 \frac{q}{\kappa} (NP)^{1/2}} \right]^{1/2}. \quad (15)$$

This mean Debye length is related by

$$\mathcal{L} = (\mathcal{L}_n \mathcal{L}_p)^{1/2} \quad (16)$$

to the n -region and p -region Debye lengths defined respectively by

$$\mathcal{L}_n = \left[\frac{kT/q}{2 \frac{q}{\kappa} N} \right]^{1/2} \quad (17)$$

and

$$\mathcal{L}_p = \left[\frac{kT/q}{2 \frac{q}{\kappa} P} \right]^{1/2}. \quad (18)$$

We now use the reference quantities defined in (13)–(15) to introduce the *normalized distance*

$$\hat{y} \equiv \frac{y}{\mathcal{L}}, \quad (19)$$

the *normalized thickness* of the intrinsic layer

$$\hat{L} \equiv \frac{L}{\mathcal{L}}, \quad (20)$$

the *normalized concentrations* of positive and negative mobile carriers

$$\hat{p} \equiv \frac{p}{(NP)^{1/2}} \quad \text{and} \quad \hat{n} \equiv \frac{n}{(NP)^{1/2}}, \quad (21, 22)$$

the *normalized potential*

$$\hat{\psi} \equiv \frac{\psi}{\psi_B}, \quad (23)$$

and the *normalized electric field intensity*

$$\hat{E} \equiv \frac{E}{\psi_B/\mathcal{L}}. \quad (24)$$

In terms of these normalized variables, the governing equations [(1')–(4')] for the intrinsic region become:

$$\frac{d\hat{E}}{d\hat{y}} = \frac{1}{2} (\hat{p} - \hat{n}), \quad (25)$$

$$\frac{d\hat{p}}{d\hat{y}} = \hat{p}\hat{E} \quad (26)$$

$$\frac{d\hat{n}}{d\hat{y}} = -\hat{n}\hat{E}, \quad (27)$$

$$\frac{d\hat{\psi}}{d\hat{y}} = -\hat{E}. \quad (28)$$

For the n -type region, (25) should be replaced by

$$\frac{d\hat{E}}{d\hat{y}} = \frac{1}{2}(\Lambda + \hat{p} - \hat{n}) \quad (25a)$$

where

$$\Lambda \equiv \left(\frac{N}{P}\right)^{1/2}. \quad (29)$$

For the p -type region, (25) should be replaced by

$$\frac{d\hat{E}}{d\hat{y}} = \frac{1}{2}(-\Lambda^{-1} + \hat{p} - \hat{n}). \quad (25b)$$

FORMAL SOLUTION OF EQUATIONS FOR INTRINSIC REGION

Division of (25)–(27) by (28) yields, after evident rearrangements of factors,

$$\frac{d\hat{E}^2}{d\hat{\psi}} = \hat{n} - \hat{p}, \quad (30)$$

$$\frac{d \ln \hat{p}}{d\hat{\psi}} = -1, \quad (31)$$

$$\frac{d \ln \hat{n}}{d\hat{\psi}} = 1. \quad (32)$$

From (31) and (32) follow

$$\hat{p} = A e^{-\hat{\psi}} \quad (33)$$

and

$$\hat{n} = A_1 e^{\hat{\psi}}, \quad (34)$$

where A and A_1 are constants of integration. The condition (11) that $\psi = 0$ for $p = n$ implies that

$$A_1 = A.$$

Substitution of (33) and (34) into (30) yields

$$\begin{aligned} \frac{d\hat{E}^2}{d\hat{\psi}} &= A(e^{\hat{\psi}} - e^{-\hat{\psi}}) \\ &= 2A \sinh \hat{\psi}. \end{aligned} \quad (35)$$

Integration of (35) now leads to

$$\hat{E} = [2A (\cosh \hat{\psi} + B)]^{1/2} \quad (36)$$

where B is another integration constant. Substitution of (36) into (28) in the form

$$\frac{d\hat{\psi}}{\hat{E}} = -d\hat{y}$$

yields after another integration

$$\int_0^{\hat{\psi}} \frac{ds}{(\cosh s + B)^{1/2}} = (2A)^{1/2}(C - \hat{y}), \quad (37)$$

where C is the fourth integration constant.

In order to express in terms of tabulated functions the relationship between $\hat{\psi}$ and $\hat{y}$ defined by (37) we shall consider two cases: $-1 < B \leq 1$ and $B \geq 1$. (It is not necessary to consider $B < -1$ because A is essentially positive [see (33)] so that $B < -1$ would imply an imaginary field strength [see (36)] at the plane in the intrinsic region where $\hat{\psi} = 0$.)

The changes of variable of integration

$$s = 2 \sinh^{-1} \cot \lambda \text{ for } -1 < B \leq 1,$$

$$s = 2 \sinh^{-1} \tan \theta \text{ for } B \geq 1$$

permit the carrying out of the integration indicated on the left side of (37). This gives

$$2^{1/2} \left(K \left[\left(\frac{1-B}{2} \right)^{1/2} \right] - F \left[\left(\frac{1-B}{2} \right)^{1/2}, \sin^{-1} \operatorname{sech} \frac{\hat{\psi}}{2} \right] \right) = (2A)^{1/2}(C - \hat{y}) \quad (38a)$$

or

$$\hat{\psi} = \ln \frac{1 + \operatorname{cn} \left(\left(\frac{1-B}{2} \right)^{1/2}, K \left[\left(\frac{1-B}{2} \right)^{1/2} \right] - A^{1/2}(C - \hat{y}) \right)}{1 - \operatorname{cn} \left(\left(\frac{1-B}{2} \right)^{1/2}, K \left[\left(\frac{1-B}{2} \right)^{1/2} \right] - A^{1/2}(C - \hat{y}) \right)} \quad (38b)$$

for $-1 < B \leq 1$

and

$$\frac{2}{(B+1)^{1/2}} F \left[\left(\frac{B-1}{B+1} \right)^{1/2}, \sin^{-1} \tanh \frac{\hat{\psi}}{2} \right] = (2A)^{1/2}(C - \hat{y}) \quad (39a)$$

or

$$\hat{\psi} = \ln \frac{1 + \operatorname{sn} \left(\left(\frac{B-1}{B+1} \right)^{1/2}, \left(\frac{B+1}{2} \right)^{1/2} A^{1/2} (C - \hat{y}) \right)}{1 - \operatorname{sn} \left(\left(\frac{B-1}{B+1} \right)^{1/2}, \left(\frac{B+1}{2} \right)^{1/2} A^{1/2} (C - \hat{y}) \right)} \quad (39b)$$

for $B \geq 1$.

Note:

$$F[k, \phi] \equiv \int_0^\phi \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}$$

is the Elliptic Integral of First Kind, usually tabulated for

$$0 \leq \phi \leq \frac{\pi}{2}, \quad 0 \leq k \leq 1.$$

$$K[k] \equiv F \left[k, \frac{\pi}{2} \right]$$

is the Complete Elliptic Integral of First Kind.

$$\operatorname{sn}[k, F] \equiv \sin \phi$$

and

$$\operatorname{cn}[k, F] \equiv \cos \phi$$

are Jacobian Elliptic Functions associated with $F[k, \phi]$.

When A , B , and C are prescribed, (38b) or (39b) gives the normalized potential $\hat{\psi}$ as a function of the normalized distance $\hat{y}$. The normalized hole concentration $\hat{p}$, electron concentration $\hat{n}$, and field intensity $\hat{E}$ are then given as functions of $\hat{y}$ by (33), (34), and (36).

In a following section it will be shown that, subject to the reasonable assumptions

$$\Lambda e^{-U/2}, \Lambda^{-1} e^{-U/2} \ll 1, \quad (40a, b)$$

$$\hat{L} \gg \Lambda^{-1/2}, \Lambda^{1/2}, \quad (40c, d)$$

$$U < \frac{1}{20} \Lambda^{1/2} L, \frac{1}{20} \Lambda^{-1/2} \hat{L}, \quad (40e, f)$$

A , C , and B are given in terms of the normalized potential at $y = 0$, the normalized thickness of the intrinsic region, and the parameter $\Lambda \equiv (N/P)^{1/2}$ by

$$A \approx \Lambda e^{-v-1}, \quad (41)$$

$$C \approx \frac{1}{2}\hat{L} + e^{1/2}(\Lambda^{1/2} - \Lambda^{-1/2}), \quad (42)$$

and

$$\Phi(B) \approx \frac{1}{2} \left(\frac{\Lambda}{e} \right)^{1/2} \hat{L} e^{-v/2} \quad (43)$$

where

$$\Phi(B) \equiv \begin{cases} K \left[\left(\frac{1-B}{2} \right)^{1/2} \right] & \text{for } -1 < B < 1 \\ \frac{\pi}{2} & \text{for } B = 1 \\ \left(\frac{2}{B+1} \right)^{1/2} K \left[\left(\frac{B-1}{B+1} \right)^{1/2} \right] & \text{for } B > 1. \end{cases}$$

Thus $\hat{\psi}$, $\hat{p}$, $\hat{n}$, and $\hat{E}$ can be computed as functions of $\hat{\psi}$ for any choice of U , $\hat{L}$, and Λ consistent with (40).

DESCRIPTION OF SAMPLE FIELD DISTRIBUTION COMPUTATIONS

The results of the foregoing analysis were used to compute $\hat{\psi}(\hat{y})$ and $\hat{E}(\hat{y})$ for $\Lambda = 1$ (symmetric case, $N = P$) and the combinations of U and $\hat{L}$ indicated by the following table:

		$\hat{L}$		
		4,000	40,000	400,000
U	7	X	X	X
	20	X	X	X
	200	X	X	X
	2,000		X	X
	20,000			X

The results of these computations are presented in Figs. 3-5 (pages 675 to 677) as plots of $\hat{\psi}/U$ versus $\hat{y}/\hat{L}$, and in Figs. 6-8 (pages 678 to 680) as plots of $\hat{E}/(2U/\hat{L})$ versus $\hat{y}/\hat{L}$. ($2U/\hat{L}$ is the average reduced field strength for the intrinsic region.) Selected curves from the above figures are replotted on non-logarithmic axes in Figs. 9 and 10 (pages 681 and 682). The curves need only be plotted for $0 \leq \hat{y}/\hat{L} \leq \frac{1}{2}$ because, by symmetry for $\Lambda = 1$, $\hat{E}(\hat{L} - \hat{y}) = \hat{E}(\hat{y})$ and $\hat{\psi}(\hat{L} - \hat{y}) = -\hat{\psi}(\hat{y})$. The ordinates and abscissas of Figures 3-10 are given in terms of the original unnormalized variables by

$$\frac{\hat{\psi}}{U} = \frac{\psi}{\psi(0)}, \quad \frac{\hat{y}}{\hat{L}} = \frac{y}{L} \quad \text{and} \quad \frac{\hat{E}}{2U/\hat{L}} = \frac{E}{2\psi(0)/L}.$$

($2\psi(0)/L$ is the average field intensity for the intrinsic region.)

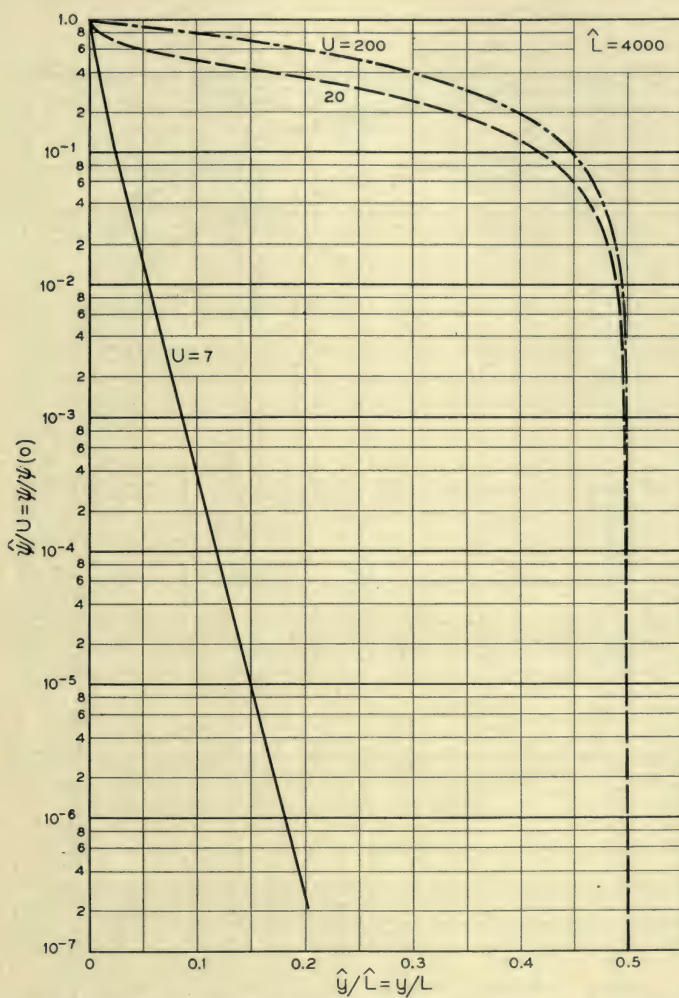


Fig. 3 — (Potential at point in intrinsic layer/Potential of *n-intrinsic* junction) versus (Distance from *n-intrinsic* junction/Intrinsic layer thickness) for $\hat{L} =$ (Intrinsic layer thickness/Mean Debye length) = 4,000 and several values of $U =$ (Potential of *n-intrinsic* junction/Boltzmann voltage).

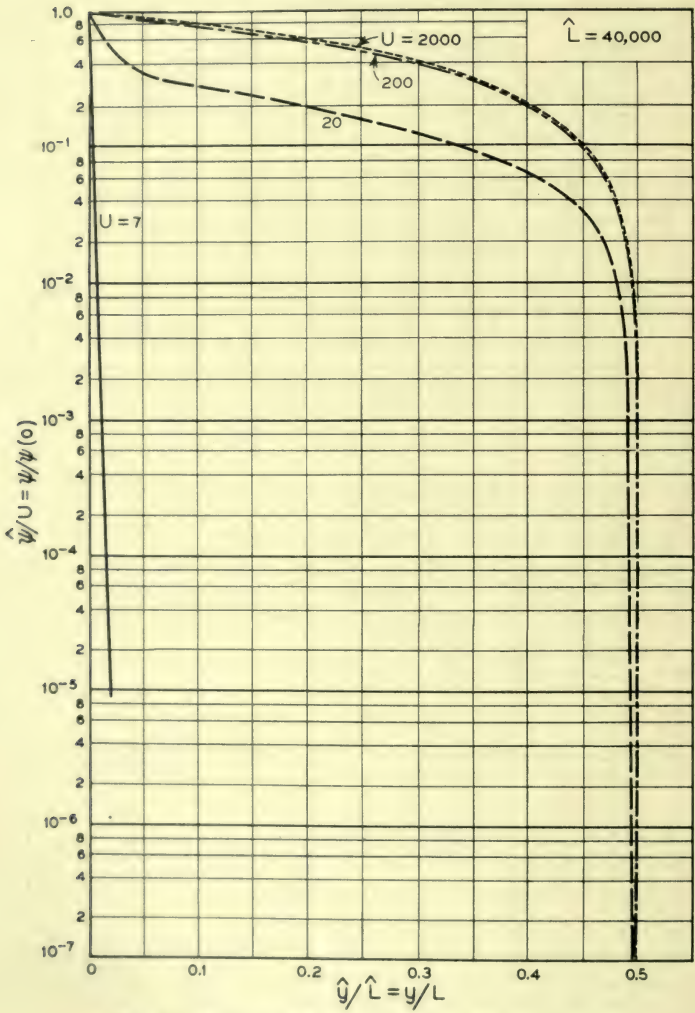
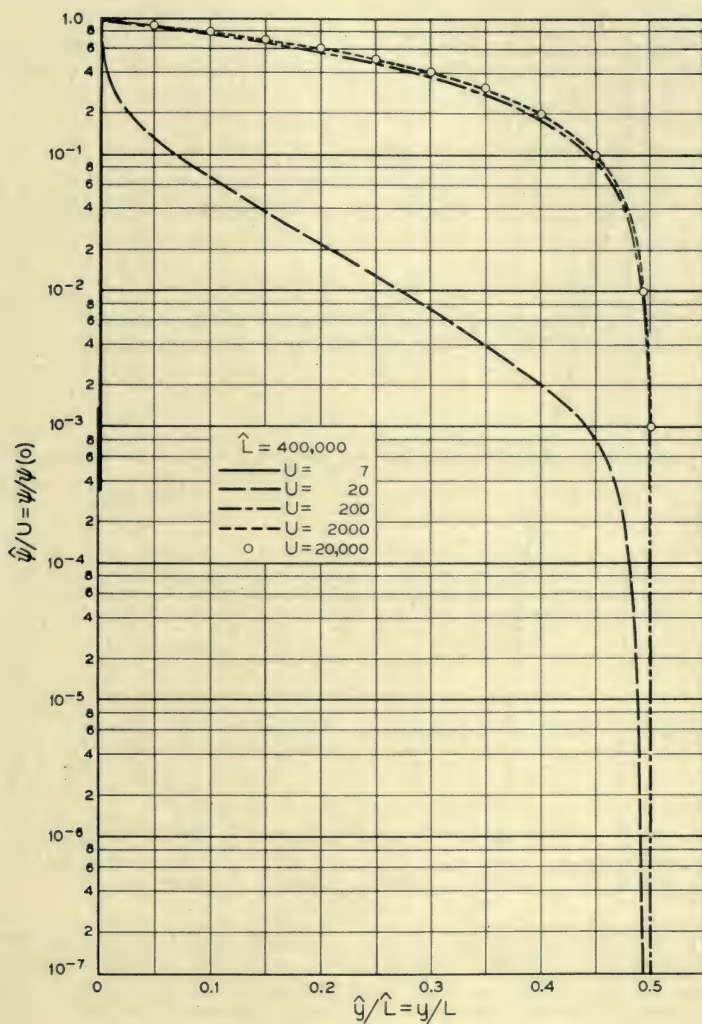


Fig. 4 — Same as Fig. 3 except for $\hat{L} = 40,000$.

Fig. 5 — Same as Fig. 3 except for $\hat{L} = 400,000$.

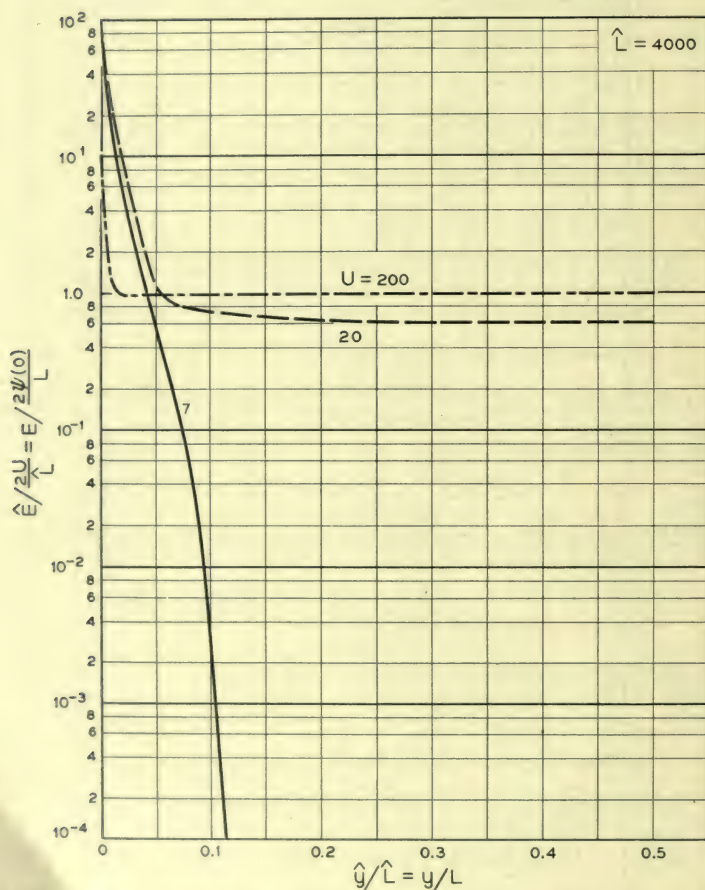
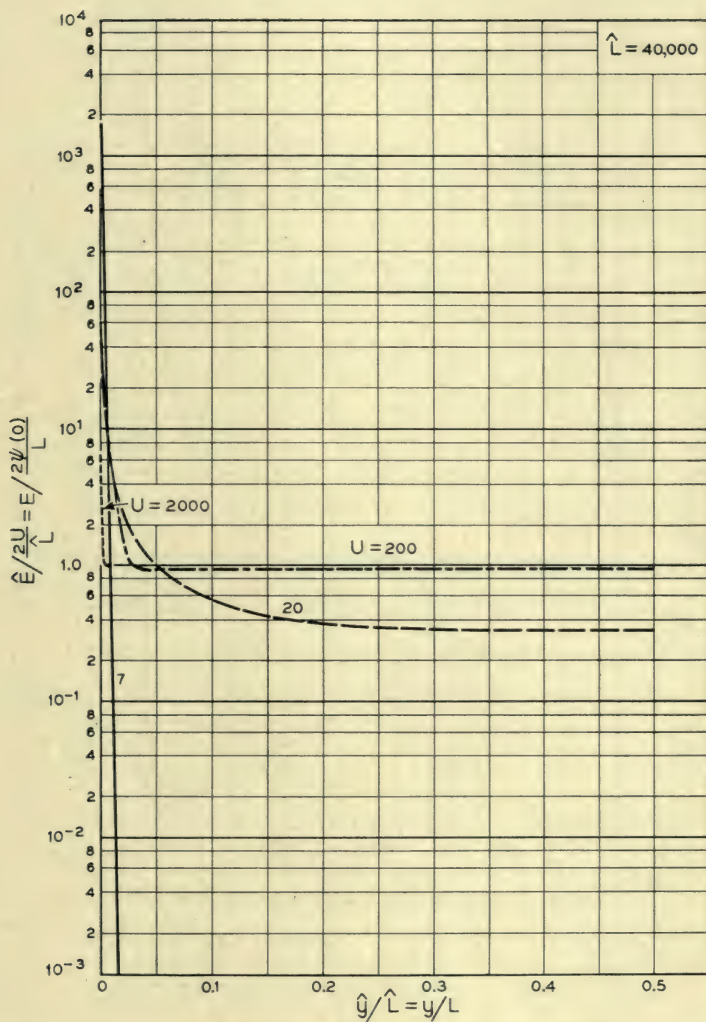
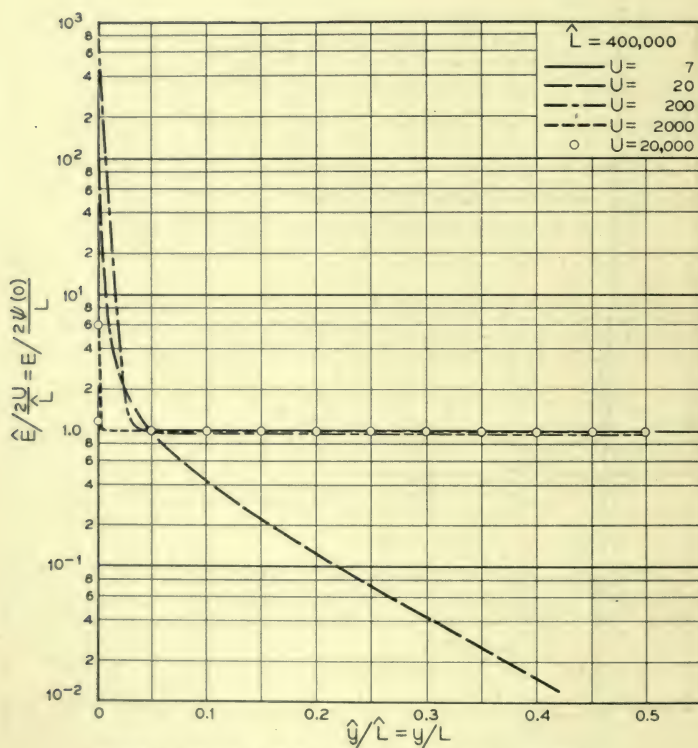


Fig. 6— (Field intensity at point in intrinsic layer/Average field intensity over intrinsic layer) versus (Distance from *n*-intrinsic junction/Intrinsic layer thickness) for $\hat{L}$ = (Intrinsic layer thickness/Mean Debye length) = 4,000 and several values of U = (Potential of *n*-intrinsic junction/Boltzmann voltage).

Fig. 7 — Same as Fig. 6 except for $\hat{L} = 40,000$.

Fig. 8 — Same as Fig. 6 except for $\hat{L} = 400,000$.

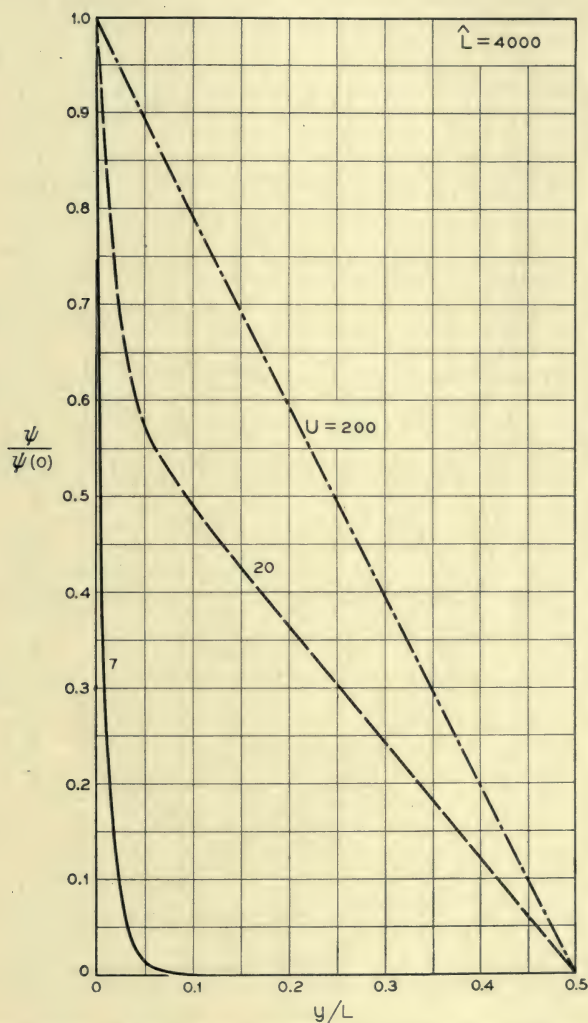


Fig. 9 — Same as Fig. 3 except with non-logarithmic potential scale.

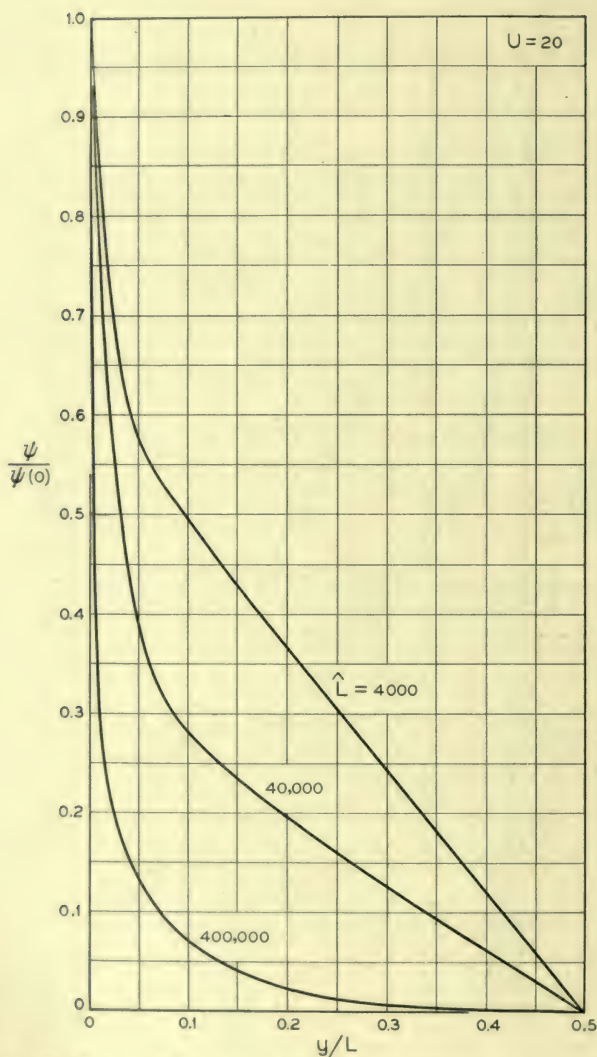


Fig. 10 — Similar to Fig. 9 except for $U =$ (Potential of n -intrinsic junction/Boltzmann voltage) = 20 and several values of $\hat{L} =$ (Intrinsic layer thickness/Mean Debye length).

Because of the extreme values of the parameters encountered in these computations, it was necessary to make extensive use of the following expansions to supplement available tables of elliptic functions:

for $k^2 \ll 1$

$$K[k] = \frac{\pi}{2} \left[1 + 2 \left(\frac{k^2}{8} \right) + 9 \left(\frac{k^2}{8} \right)^2 + \dots \right],$$

$$\operatorname{sn}[k, v] = \sin v - \frac{k^2}{4} \cos v (v - \sin v \cos v) + \dots$$

$$\operatorname{cn}[k, v] = \cos v + \frac{k^2}{4} \sin v (v - \sin v \cos v) + \dots$$

for $k' \equiv (1 - k^2)^{1/2} \ll 1$

$$K[k] = \ell n \frac{4}{k'} + \left(\ell n \frac{4}{k'} - 1 \right) \frac{k'^2}{4} + \dots$$

$$\begin{aligned} \operatorname{sn}[k, v] &= \tanh v + \frac{k'^2}{4} \operatorname{sech}^2 v (\sinh v \cosh v - v) \\ &+ \dots \end{aligned}$$

$$\begin{aligned} \operatorname{cn}[k, v] &= \operatorname{sech} v - \frac{k'^2}{4} \tanh v \operatorname{sech} v (\sinh v \cosh v - v) \\ &+ \dots \end{aligned}$$

Also useful were:

for $\phi \ll 1$

$$\ell n \frac{1 + \phi}{1 - \phi} = 2\phi \left(1 + \frac{1}{3} \phi^2 + \dots \right)$$

for $\phi \gg 1$

$$\ell n \frac{1 + \phi}{1 - \phi} = \ell n \frac{2}{1 - \phi} - \frac{1 - \phi}{2} - \frac{1}{2} \left(\frac{1 - \phi}{2} \right)^2 - \dots$$

In the determination of B from (43) the problem arises of solving the transcendental equation

$$k'K[(1 - k'^2)] = a \ll 1$$

where a is a given positive quantity. This equation must be solved by iteration or plotting and a reasonably good estimate of the root saves a great deal of labor. Making use of the approximation valid for $k' \ll 1$

$$K[(1 - k'^2)^{1/2}] \approx \ell n 4/k'$$

leads to

$$k' \ln 4/k' \approx a.$$

Now set $k' \equiv 4e^{-r}$ to obtain

$$4re^{-r} \approx a$$

or

$$f(r) \equiv r - \ln r - \ln 4/a \approx 0.$$

A rough approximation to r is obtained by neglecting $\ln r$ in comparison to r , it is $r \approx \ln 4/a$. This can be used to compute a much better approximation from Newton's formula:

$$r - r_0 \approx -\frac{f(r_0)}{f'(r_0)}.$$

The resulting useful formula is

$$r \approx \ln 4/a \left[1 + \frac{\ln \ln 4/a}{\ln 4/a - 1} \right].$$

DISCUSSION OF FIELD DISTRIBUTION CURVES

For the symmetric case ($N = P$) for which numerical computations were made, the electric field is a minimum in the middle of the intrinsic region, and rises to symmetrical maxima at the extrinsic-intrinsic interfaces. For fixed $\hat{L}$ and relatively small U the field is very small except quite near the interfaces. That is, practically all the potential drop takes place in thin "space charge layers" near the intrinsic boundaries at $\hat{y} = 0$ and $\hat{y} = \hat{L}$. As U is increased (for given $\hat{L}$) the region of appreciable field strength increases in width and the minimum field at $y = L/2$ increases. As U is made very large the minimum field strength approaches the average field over the intrinsic region. In this limit the potential distribution approaches linearity across the intrinsic region.

For $N \neq P$, the qualitative behavior of the field is the same except that the minimum field ($\hat{\psi} = 0$) moves away from $\hat{y} = \hat{L}/2$ and the maxima at $\hat{y} = 0$ and $\hat{y} = \hat{L}$ are no longer equal. This minimum field (see (38b), (39b)) occurs at

$$y = C \approx \frac{1}{2} \hat{L} + e^{1/2} (\Lambda^{1/2} - \Lambda^{-1/2})$$

Unless the asymmetry is very pronounced, this will not differ appreciably from $\hat{L}/2$ because of (40 c, d). It will be shown in a following section

that the normalized voltage at the p -intrinsic interface is given by

$$V \equiv -\hat{\psi}(\hat{L}) \approx U - 2 \ln \Lambda. \quad (44)$$

The average normalized field intensity over the intrinsic region is then given by

$$\hat{E}_{av} \equiv \frac{U + V}{\hat{L}} \approx \frac{2U}{\hat{L}} \left(1 - \frac{\ln \Lambda}{U} \right). \quad (45)$$

For $\Lambda = 1$, of course

$$\hat{E}_{av} = \frac{2U}{\hat{L}}.$$

Unless the degree of asymmetry is great,

$$\frac{\ell n \Lambda}{U}$$

will be small compared to unity and $\hat{E}_{av}$ will still be almost

$$\frac{2U}{\hat{L}}.$$

In order to present a simple quantitative picture of the penetration of the field into the interior of the intrinsic region it is convenient to introduce the *field penetration parameter* η defined by

$$\eta = \frac{\text{minimum field intensity}}{\text{average field intensity}}.$$

For fixed intrinsic region thickness and small applied voltages, $\eta \ll 1$, indicating localization of the regions of high field intensity. For large applied voltage, $\eta \rightarrow 1$, implying substantial uniformity of field throughout the intrinsic region. Subject to the assumptions (40) relations will now be derived among the applied voltage, the intrinsic layer thickness, the asymmetry parameter Λ , and the field penetration parameter η .

The minimum field intensity occurs where $\hat{\psi} = 0$. Therefore, from (36) and (45)

$$\eta \approx \frac{[2A(B + 1)]^{1/2}}{2\hat{L}^{-1}(U - \ell n \Lambda)}. \quad (46)$$

Eliminating A from (46) by the use of (41) and $\hat{L}$ by the use of (43) leads to

$$\eta \approx \frac{2^{1/2}(B + 1)^{1/2}\Phi(B)}{U - \ell n \Lambda}$$

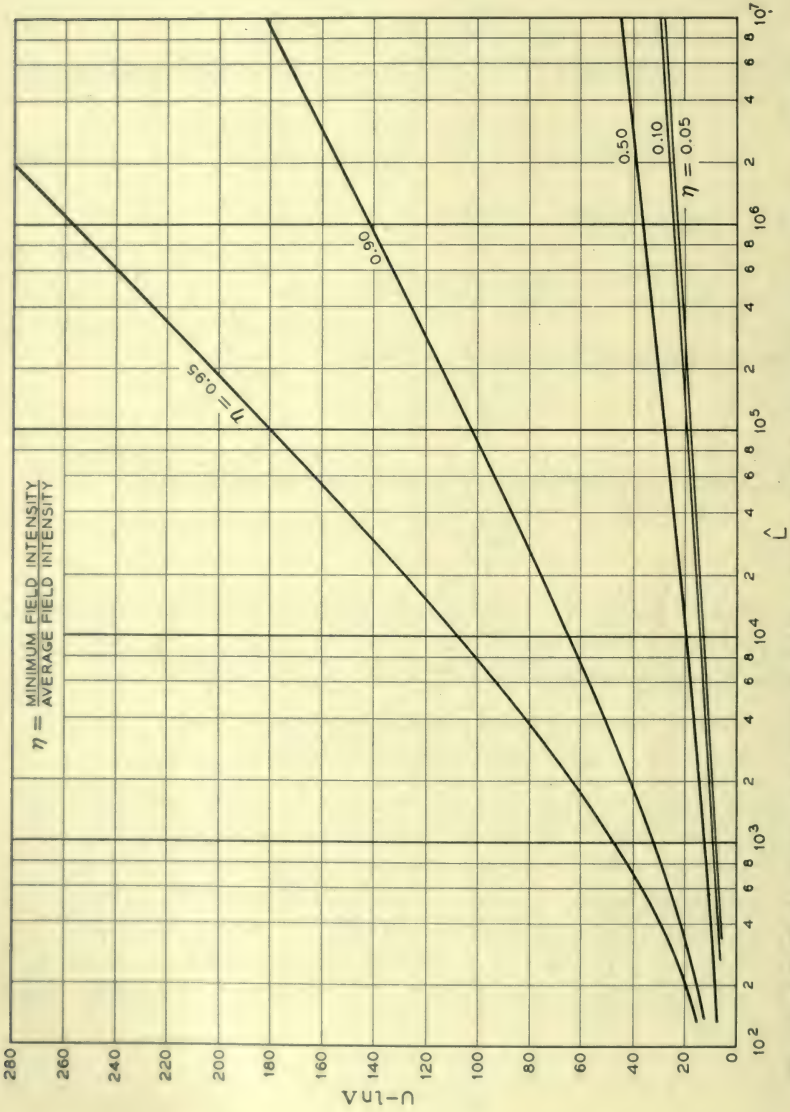


Fig. 11 — Half of voltage drop across intrinsic layer versus normalized intrinsic layer thickness for several values of the field penetration parameter η .

or

$$U - \ell n \Lambda \approx \frac{2}{\eta} \left(\frac{B+1}{2} \right)^{1/2} \Phi(B). \quad (47)$$

Then substitution of (47) into (43) yields

$$\hat{L} \approx 2e^{1/2} \Phi(B) \exp \left[\frac{1}{\eta} \left(\frac{B+1}{2} \right)^{1/2} \Phi(B) \right]. \quad (48)$$

For fixed η , (47) and (48) are parametric equations of a function $U_{\Lambda}(\hat{L})$. These equations were used to compute the curves of Fig. 11 in which $U - \ell n \Lambda$ is plotted against $\hat{L}$ for $\eta = 0.05, 0.1, 0.5, 0.9$ and 0.95 . This figure gives a quantitative picture of the dependence of the field penetration parameter η on impressed voltage and intrinsic region thickness. [The ordinate $U - \ell n \Lambda$ is $\frac{1}{2}$ the total voltage drop across the intrinsic layer in (kT/q) units.].

The foregoing analysis clarifies the progressive elimination of the low field region near the center of the intrinsic material as the applied voltage is increased for fixed L . Now the high field regions near $y = 0$ and $y = L$ will be described. Making use of (41), together with

$$\cosh U \approx \frac{1}{2} e^U$$

and

$$2AB \ll \Lambda, \Lambda^{-1}$$

implied by (40), in (36) leads to

$$\hat{E}(0) \approx e^{-1/2} \Lambda^{1/2} \quad (49)$$

and to

$$\frac{d\hat{E}}{d\hat{y}}(0) \approx -\frac{1}{2} e^{-1} \Lambda. \quad (50)$$

Hence a length characterizing the "space charge layer thickness" at $y = 0$ is

$$\frac{\hat{E}(0)}{-\frac{d\hat{E}}{d\hat{y}}(0)} \approx 2e^{1/2} \Lambda^{-1/2}. \quad (51)$$

Similarly, for the p -intrinsic interface at $\hat{y} = \hat{L}$, we have

$$E(\hat{L}) \approx e^{-1/2} \Lambda^{-1/2}, \quad (52)$$

$$\frac{d\hat{E}}{d\hat{y}}(\hat{L}) \approx -\frac{1}{2} e^{-1} \Lambda^{-1}, \quad (53)$$

and

$$\frac{\hat{E}(\hat{L})}{-\frac{d\hat{E}}{d\hat{y}}(\hat{L})} \approx 2e^{1/2}\Lambda^{1/2}. \quad (54)$$

If it is noted that the n -region and p -region Debye lengths are related to the mean Debye length $\mathfrak{L}$ and the asymmetry parameter Λ by

$$\mathfrak{L}_n = \Lambda^{-1/2}\mathfrak{L}$$

and

$$\mathfrak{L}_p = \Lambda^{1/2}\mathfrak{L}$$

it is seen that (49)–(54) can be written:

$$E(0) \approx e^{-1/2} \left[\frac{kT}{q} \text{ per } \mathfrak{L}_n \right], \quad (49')$$

$$E(L) \approx e^{-1/2} \left[\frac{kT}{q} \text{ per } \mathfrak{L}_p \right], \quad (52')$$

$$\frac{dE}{dy}(0) \approx -\frac{1}{2}e^{-1} \left[\frac{kT}{q} \text{ per } \mathfrak{L}_n \text{ per } \mathfrak{L}_n \right], \quad (50')$$

$$\frac{dE}{dy}(L) \approx -\frac{1}{2}e^{-1} \left[\frac{kT}{q} \text{ per } \mathfrak{L}_p \text{ per } \mathfrak{L}_p \right], \quad (53')$$

$$\frac{E(0)}{-\frac{dE}{dy}(0)} \approx 2e^{1/2} \quad [\text{times } \mathfrak{L}_n], \quad (51')$$

$$\frac{E(L)}{-\frac{dE}{dy}(L)} \approx 2e^{1/2} \quad [\text{times } \mathfrak{L}_p]. \quad (54')$$

Equations (49')–(54') show that the maximum field intensity and the “space charge layer thickness” at either extrinsic junction is dependent only on the Debye length of the adjacent extrinsic material.

Similarly, the concentrations of the majority carrier at either interface is found to be substantially independent of U and L , and determined by the neighboring extrinsic material:

$$\frac{n(0)}{N} \approx e^{-1}, \quad (55)$$

$$\frac{p(L)}{P} \approx e^{-1}. \quad (56)$$

The minority carrier concentration at the interfaces depends too, on the neighboring extrinsic material, but is also U dependent:

$$\frac{p(0)}{N} \approx e^{-1} e^{-2U}, \quad (57)$$

$$\frac{n(L)}{P} \approx e^{-1} e^{-2U}. \quad (58)$$

At the point of minimum field intensity, $\hat{\psi} = 0$ and

$$\frac{n}{N} = \frac{p}{N} \approx e^{-1} e^{-U}. \quad (59)$$

The extremely low carrier concentrations given by (57)–(59) are not really meaningful, of course, because the analysis has neglected the carrier concentrations due to thermally generated hole-electron pairs and to saturation currents injected through the biased junctions. While these latter concentrations are negligible compared to (55) and (56), they are undoubtedly large compared to (57) and (58). Therefore, although they can be neglected in determining the electric field intensity distribution, they are of principal importance in determining the small residue carrier concentrations in the "swept" region. A computation of these concentrations can be made by regarding the fields determined in the present analysis as impressed and studying the resulting motion of the generated and injected carriers.

APPENDIX I

EVALUATION OF INTEGRATION CONSTANTS A, B, AND C.

In this section the conditions described in (7)–(12) will be used to evaluate the integration constants A , B , and C in terms of the prescribed parameters L , U , and Λ .

First a partial integration of the differential equations for the extrinsic regions will be performed to obtain from (7)–(10) relations between $\hat{E}$ and $\hat{n}$ at $\hat{y} = 0$ and $\hat{E}$ and $\hat{p}$ at $\hat{y} = \hat{L}$. Division of (25a) by (27) gives (for the n -region $\hat{y} < 0$)

$$\frac{d\hat{E}^2}{d\hat{n}} = 1 - \frac{\hat{p}}{\hat{n}} - \frac{\Lambda}{\hat{n}}. \quad (60)$$

Addition of $\hat{n}$ times (26) to $\hat{p}$ times (27) yields

$$\frac{d}{d\hat{y}} (\hat{p}\hat{n}) = 0,$$

whence

$$\hat{p} = \frac{\nu}{\hat{n}}, \quad (61)$$

where ν is a constant. Condition (7₂) then implies

$$\nu = \frac{n_i^2}{NP}. \quad (62)$$

Substituting (61) into (60) and integrating we obtain

$$\hat{E}^2 = \hat{n} - \Lambda \ln \hat{n} + \frac{\nu}{\hat{n}} + D,$$

where D is a constant of integration. Now by (7₃), for $\hat{y} = -\infty$ $E = 0$ and

$$\begin{aligned} \hat{n} &= \Lambda + \hat{p} = \Lambda + \frac{\nu}{\hat{n}} \\ &= \Lambda \left(1 + \frac{\nu}{\Lambda \hat{n}} \right). \end{aligned}$$

Or, since

$$\begin{aligned} \frac{\nu}{\Lambda \hat{n}} &\equiv \frac{n_i^2}{nN} \approx \left(\frac{n_i}{N} \right)^2 \ll 1, \\ \hat{n} &\approx \Lambda \text{ for } \hat{E} = 0. \end{aligned}$$

Therefore

$$D \approx -\Lambda + \Lambda \ln \Lambda - \frac{\nu}{\Lambda},$$

giving

$$\hat{E}^2 \approx (\hat{n} - \Lambda) \left(1 - \frac{\nu}{\Lambda \hat{n}} \right) + \Lambda \ln \frac{\Lambda}{\hat{n}},$$

or, for $(1 - \hat{n}/\Lambda)$ positive but not $\ll 1$,

$$\hat{E}^2 \approx \hat{n} + \Lambda \ln \frac{\Lambda}{e\hat{n}} \quad [\text{for } \hat{y} = 0]. \quad (63)$$

A directly parallel computation for the p -region ($\hat{y} > \hat{L}$) leads to

$$\hat{E}^2 \approx \hat{p} + \Lambda^{-1} \ln \frac{\Lambda^{-1}}{e\hat{p}} \quad [\text{for } \hat{y} = \hat{L}]. \quad (64)$$

Upon substituting (34) and (36) into (63) and setting $\hat{\psi} = U$ (12), we obtain

$$Ae^{-U} + 2AB = \Lambda \ln \frac{\Lambda}{Ae^{U+1}}, \quad (65)$$

or, because both terms on the left are negligible under the assumptions (40)*,

$$A \approx \Lambda e^{-U-1}. \quad (66)$$

Similarly, substitution of (35) and (36) into (63) and setting of $\hat{\psi} = -V$ yields

$$Ae^{-V} + 2AB = \Lambda^{-1} \ln \frac{\Lambda^{-1}}{Ae^{V+1}}, \quad (67)$$

or, by virtue of (40),

$$A \approx \Lambda^{-1} e^{-V-1}. \quad (68)$$

Combining (66) and (68) we obtain

$$V \approx U - 2 \ln \Lambda. \quad (69)$$

This formula gives the normalized potential magnitude at the p -intrinsic interface in terms of that (U) at the n -intrinsic interface and the asymmetry parameter Λ .

Now the condition $\hat{\psi} = U$ for $\hat{y} = 0$ requires (from 38a and 39a) that

$$K \left[\left(\frac{1-B}{2} \right)^{1/2} \right] - F \left[\left(\frac{1-B}{2} \right)^{1/2} \sin_{(1)}^{-1} \operatorname{sech} \frac{U}{2} \right] = A^{1/2} C \quad (70a)$$

for $-1 < B \leq 1$,

or

$$\left(\frac{2}{B+1} \right)^{1/2} F \left[\left(\frac{B-1}{B+1} \right)^{1/2} \sin_{(1)}^{-1} \tanh \frac{U}{2} \right] = A^{1/2} C \quad (70b)$$

for $B \geq 1$.

In addition, the condition $\hat{\psi} = -V$ for $\hat{y} = \hat{L}$ requires

$$K \left[\left(\frac{1-B}{2} \right)^{1/2} \right] - F \left[\left(\frac{1-B}{2} \right)^{1/2} \sin_{(2)}^{-1} \operatorname{sech} \frac{V}{2} \right] = A^{1/2} (C - \hat{L}) \quad (71a)$$

for $-1 < B \leq 1$,

* It will be shown later that (40) implies $AB \ll 1$.

or

$$\left(\frac{2}{B+1}\right)^{1/2} F\left[\left(\frac{B-1}{B+1}\right)^{1/2}, \sin_{(4)}^{-1} \tanh\left(\frac{-V}{2}\right)\right] = A^{1/2}(C - \hat{L}) \quad (71b)$$

for $B \geq 1$.

Fortunately, because of the assumptions (40a, b) the formidable relations (70a-71b) can be simplified to

$$\Phi(B) \approx A^{1/2}C + 2e^{-U/2} \quad (72)$$

and

$$\Phi(B) \approx A^{1/2}(\hat{L} - C) + 2e^{-V/2}, \quad (73)$$

where

$$\Phi(B) \equiv \begin{cases} K \left[\left(\frac{1-B}{2} \right)^{1/2} \right] & \text{for } -1 < B < 1 \\ \frac{\pi}{2} & \text{for } B = 1 \\ \left(\frac{2}{B+1} \right)^{1/2} K \left[\left(\frac{B-1}{B+1} \right)^{1/2} \right] & \text{for } B > 1. \end{cases}$$

Subtracting (73) from (72) we obtain

$$0 = 2A^{1/2}C - A^{1/2}\hat{L} + 2e^{-U/2} - 2e^{-V/2},$$

whence, substituting (66) and (69),

$$C \approx \frac{1}{2}\hat{L} + e^{1/2}(\Lambda^{1/2} - \Lambda^{-1/2}). \quad (74)$$

Finally, substitution of (66) and (74) into (72) gives

$$\Phi(B) \approx \left[\frac{1}{2}\Lambda^{1/2}\hat{L} + e^{1/2}(\Lambda + 1) \right] e^{-U/2-1/2},$$

or, by virtue of (40c, d),

$$\Phi(B) \approx \frac{1}{2}\Lambda^{1/2}\hat{L}e^{-U/2-1/2}. \quad (75)$$

Equations (66), (69), (74), and (75) are the desired expressions for determining A , B , C , and V when values are assigned to Λ , $\hat{L}$, and U (subject to (40)). (If necessary, some or all of the restrictions (40) could

be eliminated, but the transcendental equations to be solved for A , B , C , and V would become quite formidable.)

It should be noted that (75) permits an easy determination whether the formulae for $B < 1$ or those for $B > 1$ should be used in any particular case. Since $\Phi(1) = \pi/2$,

$$B \leq 1 \text{ for } \Lambda^{1/2} \hat{L} e^{-U/2-1/2} \gtrless \pi. \quad (76)$$

APPENDIX II

CONDITIONS FOR $2AB \ll \Lambda$, Λ^{-1}

It has been stated without proof in the foregoing analysis that the conditions (40) imply $2AB \ll \Lambda$, Λ^{-1} (and hence also $AB \ll 1$). This must now be demonstrated.

For B not $\gg 1$, (40a-40d) are sufficient, for (41) shows that $A \ll 1$. However, for $B \gg 1$, the product AB is not necessarily small because of $A \ll 1$ and additional limitations are required. To establish suitable additional conditions we shall consider combinations of U , $\hat{L}$, and Λ for which AB is very small and estimate the conditions under which this smallness begins to weaken.

By eliminating U between (41) and (43) we can write

$$\Phi(B) \approx \frac{1}{2} \hat{L} A^{1/2},$$

or

$$B^{1/2} \Phi(B) \approx \frac{1}{2} \hat{L} (AB)^{1/2}. \quad (77)$$

Now for $k \approx 1$,

$$K(k) \approx \ell n \frac{4}{\sqrt{1-k^2}}.$$

Therefore, for $B \gg 1$

$$\Phi(B) = \left(\frac{2}{B+1} \right)^{1/2} K \left[\left(\frac{B-1}{B+1} \right)^{1/2} \right] \approx \frac{\ell n B}{(2B)^{1/2}}. \quad (78)$$

Substitution of (78) into (77) now yields

$$\ell n B \approx 2^{-1/2} \hat{L} (AB)^{1/2},$$

or

$$B \approx \exp [2^{-1/2} \hat{L} (AB)^{1/2}]. \quad (79)$$

From (78) and (79)

$$\Phi(B) \approx \frac{1}{2} \hat{L} (AB)^{1/2} \exp [-2^{-3/2} \hat{L} (AB)]. \quad (80)$$

Now using the expression just obtained for $\Phi(B)$ in terms of (AB) , we can eliminate $\Phi(B)$ from (43) and solve for U as a function of (AB) . After some manipulations we obtain

$$\frac{2U}{\hat{L}} \approx (2AB)^{1/2} - \frac{2(1 - \ln 2)}{\hat{L}} - \frac{2}{\hat{L}} \ln \left(\frac{2AB}{\Lambda} \right). \quad (81)$$

For $2AB\Lambda$ or $2AB\Lambda^{-1}$ of the order of 10^{-2} , (40c, d) insure that the right side of (81) will not change in order of magnitude if the two terms multiplied by $\hat{L}^{-1}$ are dropped. Therefore $2AB\Lambda$ and $2AB\Lambda^{-1}$ will be of order 10^{-2} or less if

$$\frac{2U}{\hat{L}} < \frac{1}{10} \Lambda^{1/2}, \frac{1}{10} \Lambda^{-1/2}. \quad (82)$$

Thus for a given intrinsic layer thickness there is a ceiling on the impressed voltages for which AB is negligibly small.

Transmission Properties of Laminated Clogston Type Conductors

By E. F. VAAGE

(Manuscript received December 8, 1952)

The transmission properties of ideal laminated conductors of the Clogston type are discussed by introducing the concepts of equivalent inductance, capacitance and resistance values which are analogous to their corresponding counterparts in the treatment of ordinary transmission lines. From these constants the attenuation, phase constant, and speed of propagation are obtained using conventional transmission line theory, and the results compared with those for ordinary coaxial conductors.

This paper is divided into two parts. In the first part a general discussion is given of Clogston cables and a comparison made with the conventional coaxial cable. This is illustrated with a few numerical examples, based on formulas which are developed in the second part of this paper.

INTRODUCTION

The discovery that deep penetration of the current can be obtained in laminated conductors, when the speed of propagation is made constant over the entire cross-section of the cable, is described in an earlier issue of this magazine.¹ The theoretical study of the problem was based on Maxwell's field equations dealing with a stack of parallel plates of alternate conducting and insulating layers. When applied to concentric laminated tubes, this method results in a set of extremely complex equations. S. P. Morgan has given a rigorous solution for the case when the laminated layers are of infinitesimal thickness.²

The present paper uses a different approach which leads to simpler approximate formulas. Available theoretical results are combined with simplifying approximations and certain somewhat arbitrary assumptions

¹ Clogston, A. M., Reduction of Skin Effect Losses by the use of Laminated Conductors. Bell System Tech. J., **30**, pp. 491-529, July, 1951.

² Morgan, S. P., Mathematical Theory of Laminated Transmission Lines. Bell System Tech. J., Part I, **31**, pp. 883-949, Sept., 1952, and Part 2, **31**, pp. 1121-1206, Nov., 1952.

in such a way that formulas for the counterparts of the usual transmission line properties of inductance, capacitance and resistance are obtained. The approximate formulas for attenuation, phase constant and speed of propagation are then derived using conventional transmission line theory.

Some computations of attenuation are presented which illustrate the interesting transmission properties of Clogston cables under ideal conditions. Exploratory work on this type of conductor is still in the early research stage, with some very difficult problems imposed by the need for a large number of thin layers with very close tolerances.

PART I — GENERAL DISCUSSION OF CLOGSTON CABLES AND A COMPARISON WITH CONVENTIONAL COAXIAL CABLE

1. SKIN EFFECT

An alternating current transmitted over a solid conductor has the tendency to crowd toward the surface of the wire. This phenomenon is known as the *skin effect* and the depth of penetration of the current is usually referred to as the *skin depth*. The skin depth is defined as the distance, measured from the surface toward the center of the wire, where the current density is reduced to $1/e = 0.367$. For a copper conductor it is given by:

$$\delta = \frac{2.61}{\sqrt{F_{mc}}}, \quad (1)$$

where

δ = Skin depth in mils.

F_{mc} = Frequency in megacycles.

When the skin depth is a fraction of the wire radius, the ac resistance of the wire increases about as the square root of the frequency.

Laminated conductors disclosed by Clogston have the property that the ac resistance will remain very nearly equal to the dc resistance over a wide band of frequencies, if the conducting and insulating layers can be made thin enough and sufficiently uniform. The dc resistance of a Clogston conductor will be higher than the dc resistance of a solid conductor of the same over-all dimension by a factor $(w + t)/w$, where w and t are the thicknesses of the conducting and insulating layers respectively. As discovered by Clogston, the depth of penetration in a laminated conductor is much greater than in a conductor of solid copper if the

speed of propagation is constant over the entire cross-section. A coaxial cable having an inner laminated conductor and an outer laminated sheath must obey the following relation to obtain the desired effect:

$$\mu_I \epsilon_I = \mu \epsilon \left(1 + \frac{w}{t} \right), \quad (2)$$

where:

μ_I = Permeability in space between inner and outer conductor.

μ = Permeability of laminated conductors.

ϵ_I = Dielectric constant of insulation between inner and outer conductor.

ϵ = Dielectric constant of insulating layers in the laminated conductors.

w = Thickness of copper layers.

t = Thickness of insulating layers.

In (2) the expression $\epsilon(1 + w/t)$ is of course the mean dielectric constant of the laminated conductor. Since $1/\sqrt{\mu_I \epsilon_I}$ is the speed of propagation in the main dielectric, equation (2) indicates that the speed of propagation is the same over the entire cross section of the cable. Equation (2) must be satisfied to a high degree of accuracy, otherwise deep penetration is not possible.

2. DEFINITION OF CLOGSTON CABLES

Two laminated conductors arranged as a coaxial cable are shown in Fig. 1. The inner conductor consists of a solid copper wire of diameter d_1 , over which a large number of alternate layers of insulation and copper are arranged as concentric thin tubes. The over-all diameter of the inner conductor is D_1 . The outer conductor of the coaxial cable consists of a laminated tube of inner diameter d_2 and outer diameter D_2 . The space between D_2 and d_2 is filled with thin concentric tubes of copper and insulation of the same thicknesses as for the inner conductor. The outside of the outer conductor is covered with a solid copper sheath for protection, shielding and energizing purposes. This type of cable has been named Clogston I.

By adding more layers to the outside of the inner conductor and more layers to the inside of the outer conductor, the space between them is completely filled when $d_2 = D_1$. Such a cable is shown in Fig. 2, and has been named Clogston II.

Clogston I may be thought of as a physical variant of the conventional

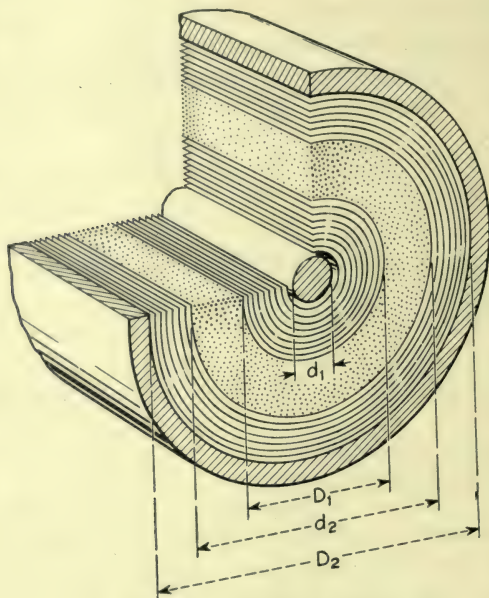


Fig. 1 — Clogston I cable.

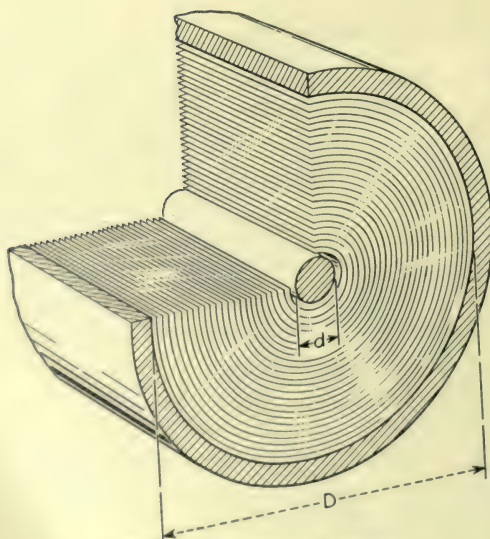


Fig. 2 — Clogston II cable.

coaxial cable, in that when one of the conductors carries a current in one direction, the other will carry a current in the opposite direction. In Clogston II there is also a reversal of currents somewhere between the outermost layers and the innermost. It is therefore a kind of a two conductor cable, but the point of division between the conductors is determined by the electromagnetic field configuration of the situation. This point has been worked out by S. P. Morgan and will be referred to in the second part of this paper.

3. OPTIMUM PROPORTIONING

The cross-sectional aspect of Clogston II is completely characterized by that proportion of the diameter D which is occupied by the laminations. This proportion is called the *Fill Factor* and is defined by:

$$\phi_{II} = (D - d)/D \quad \text{Clogston II.} \quad (3)$$

The fill factor is also a useful parameter for Clogston I, though it is not sufficient to determine its geometry. It is defined by:

$$\phi_I = (D_1 - d_1 + D_2 - d_2)/D_2 \quad \text{Clogston I.} \quad (4)$$

The additional parameters which, with the outer diameter D_2 , will completely determine the geometry, are the ratio of the over-all thickness of the inner laminated conductor to the over-all thickness of the outer conductor, and a parameter which locates the inner diameter d_1 of the inner conductor. These parameters are defined by:

$$\begin{aligned} T &= (D_1 - d_1)/(D_2 - d_2), \\ U &= d_1/D_2. \end{aligned} \quad (5)$$

In a conventional coaxial cable, shown in Fig. 3, the optimum value of attenuation³ is obtained when $D/d = 3.59$. In Clogston cables no such optimum values exist.

S. P. Morgan, however, has shown that there are useful relative optimum relations in Clogston I, which direct the choice of T and U for the cables which are illustrated in this paper. For example, for a fill factor of one-half, there is a broad optimum of attenuation when $T = 1.96$ and $U = 0.0842$. Thus the over-all thickness of the inner conductor is about twice that of the outer conductor, and the diameter d_1 of the inner core is about one-twelfth of the outer diameter D_2 . With

³ Green, E. I., F. A. Leibe and H. E. Curtis, The Proportioning of Shielded Circuits for Minimum High-Frequency Attenuation, Bell System Tech. J., **15**, pp. 248-283, April, 1936.

these dimensions the cross sectional area of the outer conductor is nearly twice that of the inner conductor, so that the optimum arises from other causes than matching the conductivity of the two conductors.

The problem of determining the relative optimum values of Clogston I, in general, is complicated, but numerical studies indicate that for values of the fill factor other than one-half, the values for T and U are not greatly different.

Another factor in Clogston cables which can be optimized is the ratio of the layer thicknesses of the conducting and the insulating layers. Clogston⁴ has shown that in the frequency range where the attenuation is substantially flat with frequency, this optimum ratio is equal to:

$$w/t = 2. \quad (6)$$

For this condition, the dc resistance of the laminated conductor is increased by $(w + t)/w$ or $3/2$ over a solid conductor. The dielectric constant of the main insulation, according to (2) above must equal $\epsilon_I = 3\epsilon$, which reduces the speed of propagation by $\sqrt{3}$, assuming $\mu_I = \mu$.

At frequencies where the attenuation begins to increase, other optimum values of w/t can be obtained, and the ratio will depend upon what top frequency is considered.

In a practical case a fill factor of unity will probably not be used. A little space in the center will be made available for a solid conductor for energizing or other purposes.

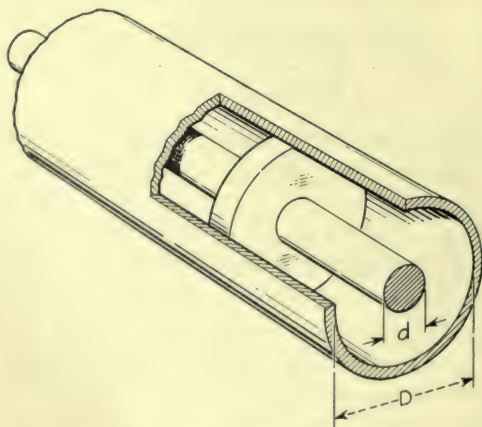


Fig. 3 — Conventional coaxial cable.

⁴ Loc. cit.

4. ATTENUATION

The attenuation of a transmission circuit at high frequencies, where $\omega L \gg R$ and $\omega C \gg G$, is usually given in the following form:

$$a = \frac{R}{2} \sqrt{\frac{C}{L}} + \frac{G}{2} \sqrt{\frac{L}{C}}, \quad (7)$$

where R is the total ac resistance of both conductors, L , C and G the inductance, capacitance and leakance of the circuit. It will be assumed that the insulation consists of polyethylene or some other material having a very low leakance. Thus as a first approximation the second term in (7) can be neglected.

In a conventional coaxial cable, in the frequency range considered, R will increase in proportion to the square root of frequency, so that neglecting leakance, (7) may be written as follows:

$$a = \frac{K_1}{D} \sqrt{F_{mc}}, \quad (8)$$

where K_1 is a constant depending upon the dielectric constant of the insulating material and the resistivity of the conductors. D is the inside diameter of the sheath, and F_{mc} the frequency in megacycles.

In the second part of this paper it is shown that the attenuation of Clogston I or II cables can be written in the following form:

$$a_c = \frac{K_2}{D_2} + K_3 w^2 F_{mc}^2, \quad (9)$$

where D = Over-all diameter of laminated cable.

w = Copper layer thickness.

K_2 and K_3 are constants, different for Clogston I and Clogston II, which depend upon the geometry of the cables, the dielectric constant of the insulating material and the resistivity of the conducting layers.

The first term in (9) gives a constant loss independent of frequency. The second term contributes little to the attenuation provided w is small enough. In fact, the attenuation will remain constant within $p\%$ of the first term provided:

$$w \leq \frac{1}{10} \sqrt{\frac{pK_2}{K_3}} \frac{1}{DF'_{mc}}. \quad (10)$$

This equation (10) determines the copper layer thickness, which will result in a "flat" attenuation within p per cent up to a frequency F'_{mc} .

By neglecting the second term in (9) and comparing the result with (8) it can be seen that the attenuation of a conventional coaxial cable and a Clogston cable are equal at a frequency given by:

$$F''_{mc} = \left[\frac{K_2}{K_1 D} \right]^2 \quad (11)$$

The numerical examples given later in this paper indicate that a Clogston cable will have higher attenuation than a conventional coaxial cable at frequencies below F''_{mc} , and less attenuation at higher frequencies. At frequencies sufficiently higher than F''_{mc} , the attenuation of a Clogston cable will increase rapidly which is evident from the second term in (9). It is in the region between F''_{mc} and frequencies where the second term in (9) becomes important that Clogston cables can theoretically provide less attenuation than a conventional coaxial cable.

5. IMPEDANCE, PHASE CONSTANT AND SPEED OF PROPAGATION

The equivalent impedances of Clogston I and Clogston II cables are developed in the second part of this paper and are equal to:

$$\begin{aligned} Z_I &= \left[1 + \frac{L_{in}}{L_{ex}} \right] \sqrt{\frac{L_{ex}}{C_{ex}}} && \text{Clogston I,} \\ Z_{II} &= \frac{L_{in}}{\sqrt{L_{ex} C_{ex}}} && \text{Clogston II.} \end{aligned} \quad (12)$$

In these expressions L_{in} , L_{ex} and C_{ex} are the internal and external inductances and capacitances respectively. For conventional coaxial cables they are discussed by S. A. Schelkunoff in "Electromagnetic Waves" (Van Nostrand, 1943). For the Clogston analogy, Part II of the present paper gives the reasoning adopted in defining them. In a Clogston II cable the external inductance goes to zero and the external capacitance to infinity, but the product $L_{ex} C_{ex}$ nevertheless remains constant.

The impedance of a conventional coaxial cable, in the frequency range considered, is given by:

$$Z = \sqrt{\frac{L}{C}} \quad (13)$$

where L and C are the external inductance and capacitance of the conventional coaxial cable.

The equivalent impedance of a Clogston cable is lower than that of conventional coaxial cable of the same outer diameter. Numerical eval-

uations using the above formulas indicate an impedance of about 23 ohms for a half-filled Clogston I cable, and about $11\frac{1}{2}$ ohms for a completely filled Clogston II cable, assuming polyethylene insulation with ϵ equal to 2.3 and copper conducting layers having a thickness twice that of the insulating layers ($w/t = 2$). These values compare with about 76 ohms for a conventional coaxial cable having air dielectric and the same outer diameter and 51 ohms for a corresponding coaxial with solid polyethylene dielectric.

The phase constants of both Clogston cables in the frequency range where the attenuation is nearly flat, are equal and independent of the geometry of the cables. They are given by:

$$\beta_I = \beta_{II} = \omega \sqrt{L_{ex} C_{ex}}, \quad (14)$$

assuming uniform layer thicknesses.

For a conventional coaxial cable the phase constant, neglecting leakage, is given approximately by:

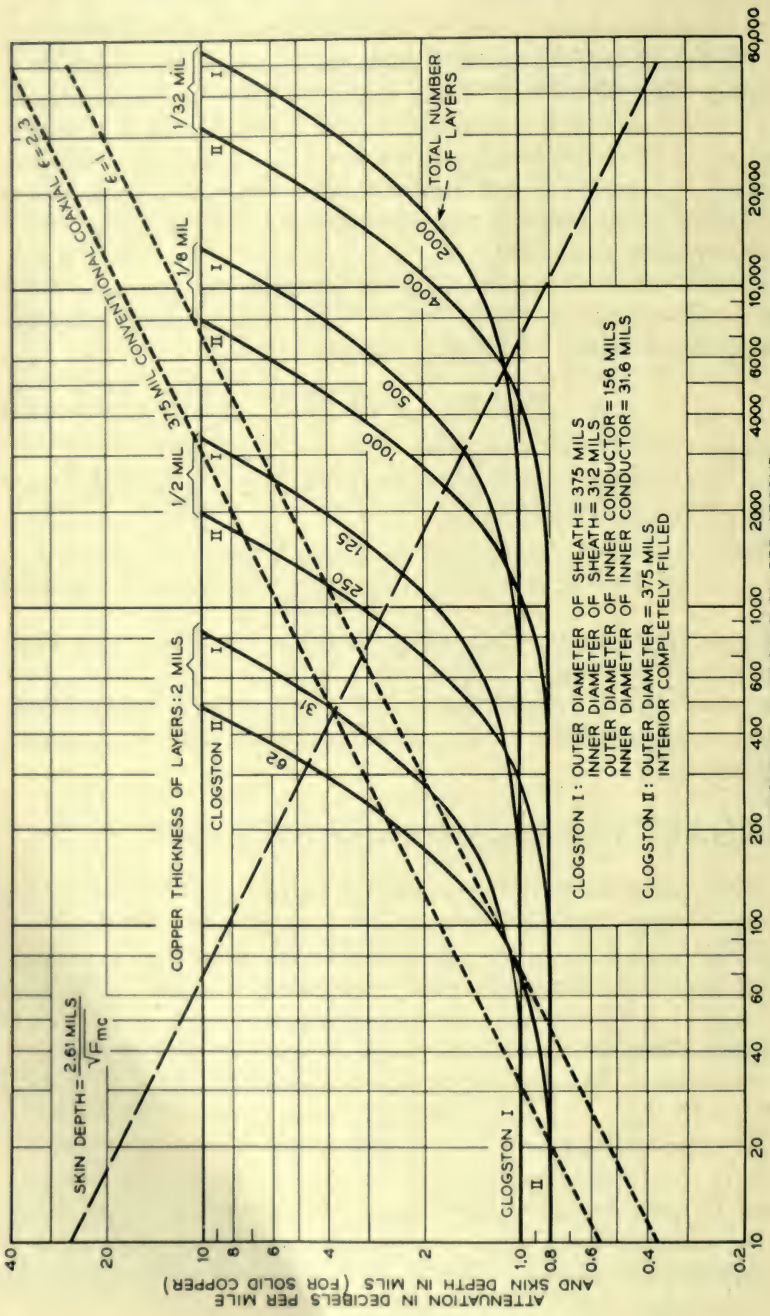
$$\beta = \omega \sqrt{LC} \left[1 - \frac{1}{2} \left(\frac{R}{2\omega L} \right)^2 \right]. \quad (15)$$

The computed speed of propagation, which is equal to ω/β , is about 71,000 mi/sec for Clogston cables with polyethylene insulating layers. This compares with 123,000 mi/sec for a conventional coaxial cable with polyethylene insulation and 186,000 mi/sec for a coaxial with pure air dielectric.

6. COMPARISON WITH A CONVENTIONAL COAXIAL CABLE

To illustrate the effect of the various parameters involved in the attenuation of a Clogston cable, and to compare the result with a conventional coaxial cable, a few numerical examples have been evaluated.

A one-half filled Clogston I and a completely filled Clogston II have been selected arbitrarily for comparison purposes. Fig. 4 shows the attenuation characteristics of these cables for several values of copper layer thicknesses. In each case, polyethylene insulation ($\epsilon = 2.3$) is assumed, with $w/t = 2$, i.e., the insulating layers have one-half the thickness of the conducting layers. In the same figure, the attenuation characteristics of two conventional coaxial cables of the same outer diameter, one with air dielectric and one with polyethylene insulation, are shown also. The regions where Clogston cables have in theory less attenuation than conventional coaxial cables of the same outer diameter can be seen in this figure.



The dotted curve shown on Fig. 4 gives the skin depth in solid copper. It will be noted that the copper layer thicknesses become smaller and smaller fractions of the skin depth as the frequency increases. This is also evident from (1) and (10) above, which show that the copper layer thicknesses are inversely proportional to the frequency, while the skin depth is inversely proportional to the square root of the frequency.

The effect of fill is illustrated in Fig. 5 for a 375-mil Clogston I cable, for fill factors of one-eighth, one-quarter, and one-half. It will be noted that the attenuation increases rapidly with decrease in fill, accompanied by an increase in the frequency band over which the attenuation is flat. The attenuation of a completely filled Clogston II cable is also shown for comparison.

The above estimates are based on ideal conditions; that is, it is assumed that the laminated structures are perfectly uniform. The effects of departures from ideal are not shown by the approximate methods used in this paper, but it has been shown by Morgan⁵ that even small departures from ideal conditions will result in increases and irregularities in attenuation, and a decrease in the band over which the attenuation is approximately uniform.

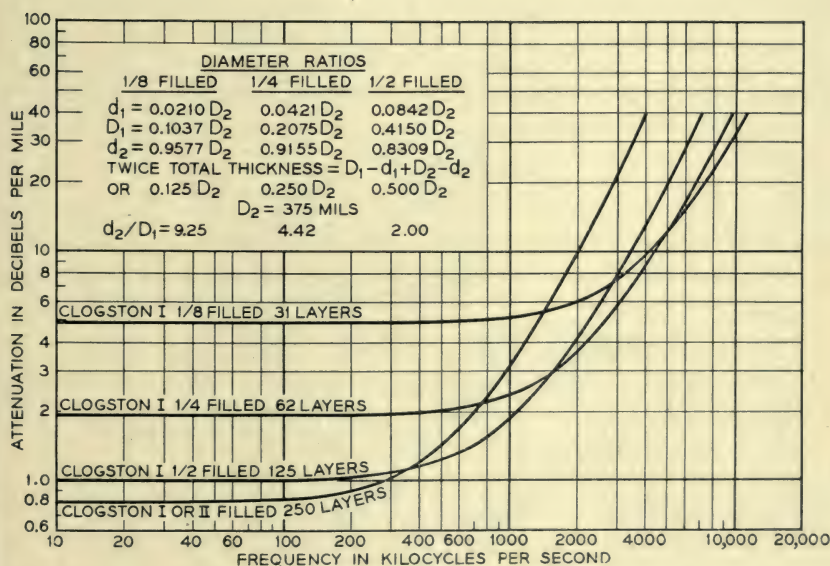


Fig. 5 — Attenuation of Clogston I.

⁵ Loc. cit. Part II, pages 1161-1201.

PART II — DERIVATION AND EVALUATION OF FORMULAS

In this part of the paper, approximate formulas for the counterparts of the usual primary constants used in transmission line computations are derived for Clogston I and II type cables. From these formulas the various constants entering into the expressions for attenuation, impedance and phase as given in Part I are evaluated in terms of the dimensions of the cables and the frequency.

In deriving the formulas, certain simplifying assumptions have been made as explained in the test. The effects of these assumptions are examined by comparing the results with those obtained by more rigorous methods.

1. RESISTANCE IN GENERAL

A Clogston I cable consists of a laminated inner conductor and a laminated outer conductor as shown in Fig. 1. Each of these may be represented by the laminated conductor shown in Fig. 6. At present, we have no exact formula for the ac resistance of such conductors over a wide range of frequencies. However, S. P. Morgan has shown that for a stack of parallel plates, the resistance per unit cross-sectional area is equal to:

$$R_{ac} = R_{dc} \frac{w + t}{w} \left[1 + \frac{(n^2 - 1/5)w^4}{412} F_{mc}^2 + \dots \right], \quad (16)$$

where R_{dc} is the dc resistance of the stack, when completely filled with

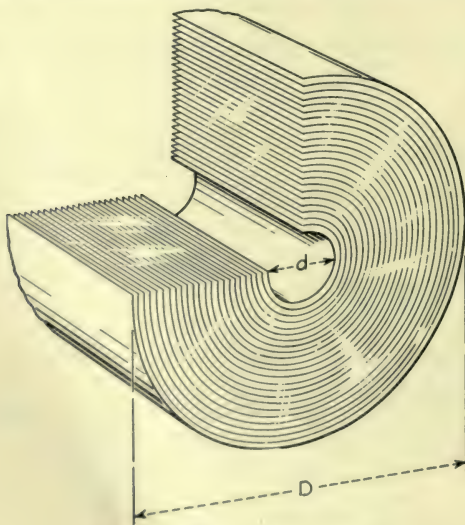


Fig. 6 — Laminated conductor.

copper. The other parameters are:

w = Thickness of copper layers in mils.

t = Thickness of insulating layers in mils.

n = Total number of layers.

F_{mc} = Frequency in megacycles.

With curvature disregarded, equation (16) also gives the ac resistance of the laminated conductor shown in Fig. 6. For only one copper layer, $n = 1$, and $t = 0$, (16) reduces to:

$$R_{ac} = R_{dc} \left[1 + \frac{w^4 F_{mc}^2}{515} + \dots \right], \quad (17)$$

which is exactly the expression for a copper tube at frequencies where the ac resistance begins to depart from the dc resistance and $\frac{1}{2}(D - d) = w$. For a single layer, the effect of curvature is very small⁶ and can be disregarded. It will be assumed that (16) also holds to a fair degree of accuracy for a laminated conductor made up of a large number of layers. Since n is large, the small fraction one-fifth can be neglected. For the optimum condition (minimum attenuation) Clogston⁷ has shown that:

$$w = 2t. \quad (18)$$

The total number of layers can be obtained from the following expression:

$$n = \frac{D - d}{2(w + t)} = \frac{D - d}{3w} \text{ for } w = 2t. \quad (19)$$

With (18) and (19) substituted in (16) the ac resistance is given by:

$$R_{ac} = \frac{82080}{D^2 - d^2} \left[1 + \frac{w^2(D - d)^2 F_{mc}^2}{3710} + \dots \right], \quad (20)$$

where the diameters and the copper layer thickness are given in mils and the frequency in megacycles. The resistivity of copper is taken to be 1.724×10^{-6} ohms/cm³.

2. CLOGSTON I CABLE

2.1 Resistance

The ac resistances of the inner and outer laminated conductors of Clogston I cable, shown in Fig. 1, can be obtained from (20) above by substituting D_1 and d_1 for the inner conductor and D_2 and d_2 for the

⁶ Schelkunoff, S. A., The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields. Bell System Tech. J., **13**, pp. 532-579, Oct., 1934. BSTJ, October, 1934, by S. A. Schelkunoff.

⁷ Loc. cit.

outer conductor and adding the result. Thus:

$$R_{ac1} + R_{ac2} = \frac{82080A_1}{D_2^2} \left[1 + \frac{B_1 w^2 D_2^2 F_{mc}^2}{3710} + \dots \right] \text{ ohms/mi,} \quad (21)$$

where:

$$A_1 = \frac{D_2^2}{D_1^2 - d_1^2} + \frac{D_2^2}{D_2^2 - d_2^2}, \quad (22)$$

$$B_1 = \frac{2(D_1 D_2 - d_1 d_2)(D_1 - d_1)(D_2 - d_2)}{(D_1^2 - d_1^2 + D_2^2 - d_2^2)D_2^2}. \quad (23)$$

From (4) and (5) in the first part of this paper, it is possible to express A_1 and B_1 wholly in terms of ϕ_I , T and U . Thus, A_1 and B_1 are independent of D_2 , and it follows that the first term in (21) is inversely proportional to D_2^2 , while the second term, when multiplied out, is independent of D_2 , assuming fixed values of ϕ_I , T and U .

2.2 Impedance, Inductance and Capacitance

In a coaxial cable the flux in the space between the two conductors gives the external inductance. The internal inductance is obtained from considering the flux within the walls of the conductors themselves but not that in the space between them. The effective inductance is then the sum of the two. Analogous considerations apply to the external and internal capacitance and the effective over-all capacitance is the value of the two acting in series.

Similarly in the frequency range where $\omega L \gg R$ and $\omega C \gg G$, but where the ac resistance is nearly equal to the dc resistance, the internal inductance and capacitance of the laminated conductors must be taken into account. Since they are in series with the external components they tend to increase the total inductance and decrease the total capacitance. The impedance of the circuit can therefore be expressed as follows:

$$Z_I = \sqrt{\frac{L}{C}} = \sqrt{\frac{L_{in} + L_{ex}}{\left[\frac{C_{in} C_{ex}}{C_{in} + C_{ex}} \right]}} \text{ ohms,} \quad (24)$$

$$L_{ex} = \frac{\mu_I}{2\pi} \ell n \left(\frac{d_2}{D_1} \right) \text{ henries/cm,} \quad (25)$$

$$C_{ex} = \frac{2\pi\epsilon_I}{\ell n \left(\frac{d_2}{D_1} \right)} \text{ farads/cm,} \quad (26)$$

where: $\epsilon_I = \epsilon(1 + w/t)$ = Dielectric constant of insulation between inner and outer conductors.

At the present time no exact formulas for the internal components of either L or C are available. The internal inductance, however, must be nearly equal to that of a solid wire for the inner laminated conductor and to that of a solid sheath for the outer laminated conductor, and will be assumed so in this paper.

It should be remembered that deep penetration of the currents will be obtained when Clogston condition of constant speed of propagation over the entire cross section of the cable is satisfied. The speed of propagation over the laminated conductors is equal to $1/\sqrt{L_{in}C_{in}}$ and in the main dielectric between the two conductors equal to $1/\sqrt{L_{ex}C_{ex}}$. Thus to obtain the Clogston condition the following relation must hold:

$$L_{in}C_{in} = L_{ex}C_{ex}. \quad (27)$$

By solving for the unknown quantity C_{in} and substituting the result in (24), the impedance of Clogston I cable is found to be:

$$Z_I = \left[1 + \frac{L_{in}}{L_{ex}} \right] \sqrt{\frac{L_{ex}}{C_{ex}}} \text{ ohms.} \quad (28)$$

Formulas for L_{in} , L_{ex} and C_{ex} in units convenient for numerical evaluation are given later in the section summarizing the formulas.

2.3 Attenuation and Phase

The attenuation of Clogston I cable neglecting leakance is obtained from the following expression:

$$\alpha_I = \frac{R_{ac1} + R_{ac2}}{2Z_I}. \quad (29)$$

The phase constant, at the frequencies considered, and where the ac resistance of the conductors does not depart appreciably from the dc value, is equal to:

$$\beta_I = \omega \sqrt{LC} = \omega \sqrt{(L_{in} + L_{ex}) \frac{C_{in}C_{ex}}{C_{in} + C_{ex}}}, \quad (30)$$

which with (27) above reduces to:

$$\beta_I = \omega \sqrt{L_{ex}C_{ex}}. \quad (31)$$

3. CLOGSTON II CABLE

3.1 Resistance

A Clogston II cable may be looked upon as having an "inner" and "outer" conductor which are separated by an infinitesimal amount in which R_{ac} given by (20) above is an expression for the parallel connection of R_{ac1} and R_{ac2} of a Clogston II cable having an outer diameter of D and an inner diameter of d . That is:

$$R_{ac} = \frac{R_{ac1}R_{ac2}}{R_{ac1} + R_{ac2}}. \quad (32)$$

It will now be assumed that (1) the currents flowing in one direction through the "inner" conductor and in the opposite direction in the "outer" conductor, will separate in such a way that R_{ac1} is equal to R_{ac2} , and (2) that the currents are uniformly distributed over the cross sections of the conductors. With these assumptions the respective cross sections would then be equal, and the reversal of the current would take place in a completely filled ($d = 0$) Clogston II cable at a radius equal to:

$$r = \frac{D}{2\sqrt{2}} = 0.3535D. \quad (33)$$

By substituting (32) and (33) in (20) it can be shown that:

$$R_{ac1} + R_{ac2} = \frac{328320}{D^2 - d^2} \left[1 + \frac{B_2 w^2 D^2 F_{mc}^2}{3710} + \dots \right] \text{ ohms/mi}, \quad (34)$$

where

$$B_2 = 1 + \frac{d^2}{D^2} - \left[1 + \frac{d}{D} \right] \sqrt{\frac{1}{2} \left[1 + \frac{d^2}{D^2} \right]}. \quad (35)$$

The above assumptions relating to division and distribution of the current are not exact. S. P. Morgan has shown that the current distribution is not uniform and that the reversal of the current takes place at a radius equal to $0.3138D$. As shown in a later section of this paper, the error resulting from these simplifying assumptions is not large.

3.2 Impedance, Attenuation and Phase

In a Clogston II cable, the main dielectric insulation between "inner" and "outer" conductor has vanished. Thus the external inductance approaches zero and the external capacitance becomes infinitely large.

From (25) and (26) above it is evident, however, that the product of $L_{\text{ex}}C_{\text{ex}}$ remains constant, since it is independent of the diameter ratio d_2/D_1 , which in a Clogston II cable approaches unity. With (27) inserted in (24) and with $L_{\text{ex}} = 0$ and $C_{\text{ex}} = \infty$, the impedance of a Clogston II cable may be written:

$$Z_{\text{II}} = \sqrt{\frac{L_{\text{in}}}{C_{\text{in}}}} = \frac{L_{\text{in}}}{\sqrt{L_{\text{ex}}C_{\text{ex}}}} \text{ ohms.} \quad (36)$$

The internal inductance of a Clogston II cable is not known, but will be taken equal to 0.1609×10^{-3} Henries/mi, which is the internal inductance of a pair of wires at low frequency. Thus:

$$Z_{\text{II}} = \frac{17.35}{\sqrt{\epsilon}} \text{ ohms,} \quad (37)$$

where ϵ is the dielectric constant of the insulating layers.

The attenuation is obtained by dividing $R_{\text{ac1}} + R_{\text{ac2}}$ from (34) by $2Z_{\text{II}}$, where leakance is disregarded.

The phase constant, in the frequency range considered, is equal to the phase constant of a Clogston I cable since $L_{\text{ex}}C_{\text{ex}}$ is a constant value.

4. Comparison of Results with Values Obtained from Rigorous Formulas

S. P. Morgan⁸ has developed rigorous formulas for the attenuation of Clogston cables, assuming infinitesimal thickness of the layers but retaining a fixed ratio of copper to insulating layer thicknesses. A correction term gives the increase in attenuation with frequency for layers of finite thickness.

The attenuation of a one-half filled Clogston I cable computed by the approximate formulas given in the present paper was 1.1 per cent higher than the value computed using Morgan's rigorous formulas. Similar computations on a completely filled Clogston II cable gave values 8.6 per cent higher. This decrease in accuracy with increase in fill is in line with the expectation that uniform distribution of the current is more closely approximated with low percentages of fill.

5. SUMMARY OF FORMULAS

The formulas developed in the second part of this paper and those for which the derivation has been indicated are summarized below. Conducting layers of copper, and insulating layers with a dielectric constant

⁸ Loc. cit.

of ϵ and a thickness half that of the conducting layers ($w/t = 2$) are assumed throughout.

5.1 Clogston I Cable

$$R_I = \frac{82080A}{D_2^2} \left[1 + \frac{B_1 w^2 D_2^2 F_{mc}^2}{3710} + \dots \right] \text{ ohms/mile,} \quad (38)$$

$$L_{in} = 0.741 \times 10^{-3} \frac{D_2^2}{D_2^2 - d_2^2} \left[\frac{D_2^2}{D_2^2 - d_2^2} \log_{10} \frac{D_2}{d_2} - 0.2172 \right] \text{ henries/mile,} \quad (39)$$

$$L_{ex} = 0.741 \times 10^{-3} \log_{10} \frac{d_2}{D_1} \text{ henries/mile,} \quad (40)$$

$$C_{ex} = \frac{0.0388 \times 10^{-6} \epsilon_I}{\log_{10} \frac{d_2}{D_1}} \text{ farads/mile,} \quad (41)$$

$$\epsilon_I = \epsilon(1 + w/t), \quad (42)$$

$$Z_I = \left[1 + \frac{L_{in}}{L_{ex}} \right] \sqrt{\frac{L_{ex}}{C_{ex}}} \text{ ohms,} \quad (43)$$

$$\beta_I = \omega \sqrt{L_{ex} C_{ex}} \text{ radians/mile,} \quad (44)$$

$$\alpha_I = \frac{K_2}{D_2^2} + K_3 w^2 F_{mc}^2 \text{ nepers/mile,} \quad (45)$$

where:

$$K_2 = \frac{41040}{Z_I} \left[\frac{D_2^2}{D_1^2 - d_1^2} + \frac{D_2^2}{D_2^2 - d_2^2} \right], \quad (46)$$

$$K_3 = \frac{22.124}{Z_I} \frac{D_1 D_2 - d_1 d_2}{(D_1 + d_1)(D_2 + d_2)}. \quad (47)$$

5.2 Clogston II Cable

$$R_{II} = \frac{328320}{D^2 - d^2} \left[1 + \frac{B_2 w^2 D^2 F_{mc}^2}{3710} + \dots \right] \text{ ohms/mile,} \quad (48)$$

$$Z_{II} = \frac{17.35}{\sqrt{\epsilon}} \text{ ohms,} \quad (49)$$

$$\beta_{II} = \omega \sqrt{L_{ex} C_{ex}} \text{ radians/mile,} \quad (50)$$

$$\alpha_{II} = \frac{K'_2}{D^2} + K'_3 w^2 F_{mc}^2 \text{ nepers/mile,} \quad (51)$$

$$K'_2 = \frac{164160}{Z_{II}} \frac{D^2}{D^2 - d^2}, \quad (52)$$

$$K'_3 = \frac{44.248}{Z_{II}} \frac{D^2 + d^2 - (D + d)\sqrt{\frac{1}{2}(D^2 + d^2)}}{D^2 - d^2}, \quad (53)$$

$$A_1 = \frac{D_2^2}{D_1^2 - d_1^2} + \frac{D_2^2}{D_2^2 - d_2^2}, \quad (54)$$

$$B_1 = \frac{2(D_1 D_2 - d_1 d_2)(D_1 - d_1)(D_2 - d_2)}{(D_1^2 - d_1^2 + D_2^2 - d_2^2)D_2^2}, \quad (55)$$

$$B_2 = 1 + \frac{d^2}{D^2} - \left(1 + \frac{d}{D}\right) \sqrt{\frac{1}{2}\left(1 + \frac{d^2}{D^2}\right)}. \quad (56)$$

The parameters in (38) to (56) are defined as follows:

D_2 = Outside diameter of outer laminated conductor in a Clogston I cable, in mils.

d_2 = Inside diameter of outer laminated conductor in a Clogston I cable, in mils.

D_1 = Outside diameter of inner laminated conductor in a Clogston I cable, in mils.

d_1 = Inside diameter of inner laminated conductor in a Clogston I cable, in mils.

D = Outside diameter of a laminated Clogston II cable, in mils.

d = Inner diameter of a laminated Clogston II cable, in mils.

w = Thickness of copper layers in mils.

F_{mc} = Frequency in megacycles.

ϵ_I = Dielectric constant of insulation between inner and outer laminated conductors in a Clogston I cable.

ϵ = Dielectric constant of insulating layers.

ACKNOWLEDGMENT

The author wishes to express his appreciation to H. S. Black, C. W. Carter, Jr., J. T. Dixon and F. B. Llewellyn for valuable assistance and advice in preparation of this paper.

A Coupled Resonator Reflex Klystron*

By E. D. REED

(Manuscript received March 5, 1953)

The theory of a coupled resonator reflex klystron is developed and its reduction to practice described. This tube differs from the conventional reflex klystron in that its performance characteristic is derived from the interaction between the electronic admittance due to a bunched electron stream and the input admittance of two synchronously tuned, coupled resonators. As a result: (1) power output can be made to be substantially flat over the greater part of the electronic tuning range; (2) the half-power electronic tuning range of the coupled resonator reflex klystron is more than twice that of a klystron using the same electron optical system but interacting with a single resonator; and (3) modulation linearity may be obtained over a greatly increased frequency swing.

A reduction in power output of about 3db occurs for a secondary resonator Q and coupling coefficient adjustment designed to yield a maximum flat band or maximum electronic tuning while a much smaller reduction in output power will provide a substantial improvement in modulation linearity.

TABLE OF CONTENTS

1.0 Introduction.....	716
2.0 Small Signal Reflex Klystron Theory.....	718
2.1 Electronic Admittance.....	719
2.2 Passive Circuit Admittance of Single Resonator.....	722
2.3 Equivalent Circuit of Single Resonator Reflex Klystron....	723
2.4 Performance Analysis Based on Complex Admittance Plane Representation.....	723
3.0 Theory of Coupled Resonator Reflex Klystron.....	727
3.1 Driving Point Properties of Two Coupled Resonators Having Equal Q's.....	729
3.1.1 Variation of Input Impedance with Frequency.....	735
3.1.2 Input Admittance Plot in g - b Plane.....	736

* Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Faculty of Pure Science of Columbia University.

3.1.3	Variation of Input Phase Angle with Frequency.....	737
3.2	Mode Shapes Resulting from Interaction Between Electronic Admittance and Input Admittance of Two Coupled Resonators of Equal Q 's.....	739
3.3	Driving Point Properties of Two Coupled Resonators Having Unequal Q 's.....	742
3.4	Mode Shapes Obtainable With Two Coupled Resonators of Unequal Q 's.....	746
4.0	An Experimental Coupled-Resonator Reflex Klystron.....	746
4.1	Constructional Features of Experimental Tube and Circuit.....	748
4.2	Qualitative Verification of Theory.....	750
4.3	Quantitative Verification of Theory.....	750
4.3.1	Determination of Primary Q	752
4.3.2	Calibration of Secondary Resonator.....	754
4.3.3	Comparison of Experimental and Theoretical Mode Shapes.....	757
4.4	Performance Data.....	760
5.0	Applications of the Coupled Resonator Reflex Klystron.....	762
6.0	Conclusions.....	764
	References.....	765

1.0 INTRODUCTION

The conventional reflex klystron derives its performance characteristics from the interaction between the electronic admittance due to a bunched electron stream and the input admittance of a resonant cavity. As is well known, this interaction results in mode shapes which are closely related to certain input properties of the passive resonant circuit. Thus, the dependence of power output upon frequency, which results from variations in repeller voltage about its mid-mode value, bears close resemblance to the input-impedance-versus-frequency plot of a parallel resonant circuit. Similarly, the curve relating frequency to repeller voltage has the same general shape as that relating frequency to the input phase angle of the resonator. Recognition of these relationships has resulted in the consideration of different and, perhaps, more useful mode shapes which might be obtained if the electronic admittance were made to interact with impedance or admittance functions of passive circuits other than that due to a single resonator.

What do we mean by "more useful" mode shapes? The answer, of course, depends on the application, although an "ideal" mode shape could probably be defined as one having a flat top, i.e., power output inde-

pendent of repeller voltage, with frequency linearly related to the latter. Moreover, these conditions should preferably obtain over the widest possible frequency range. A tube possessing such characteristics would prove exceedingly useful in a large number of applications. To list a few:

- (a) Electronically swept signal generator,
- (b) FM deviator,
- (c) Transmitting oscillator in radio relay systems employing frequency modulation, and
- (d) Local oscillator in microwave receivers with wide range AFC applied to the repeller.

Inability to realize this ideal mode shape in a practical tube might make a compromise solution appear acceptable, one consisting of a reflex klystron having a variable mode shape, i.e., a characteristic which could be adjusted to fit a particular need. Thus, application (a) requires constant power output with minor emphasis on frequency-repeller-voltage linearity, whereas applications (b) and (c) demand a high degree of modulation linearity with constancy of power output of no great importance. In application (d) the emphasis is on wide electronic tuning with both variation in power output and non-linearity in the frequency characteristic permissible.

A method to obtain this variable mode shape was achieved by the use of coupled cavities. Instead of having the bunched electron stream interact with the electric field of a single resonator, as is done in the conventional reflex klystron, we can present to it the input admittance of two coupled cavities. The resultant mode shape may then be expected to resemble the input impedance of two coupled resonant circuits just as the conventional klystron mode resembles that of a single resonator. Moreover, the input impedance of two coupled cavities can assume a large variety of contours depending on the ratio of Q 's of the primary and secondary resonators and on the tightness of coupling between the two. If we were now to construct such a double cavity reflex klystron with provision to vary the secondary Q as well as the coupling coefficient continuously, we would have the means of producing a large variety of mode shapes within a single tube. Depending on the application, the characteristics could then be adjusted to give either a range of flat power output or optimum modulation linearity or wide electronic tuning.

As might be expected at this point, a price must be paid for the advantages gained in the coupled cavity approach. It consists of the power expended in supplying the losses due to the secondary resonator. This power subtracts directly from the available useful power and, there-

fore, though resulting in broadband operation, greatly improved modulation linearity and a number of other useful properties, leads to a definite reduction in output level. Whether this can be tolerated will again depend on the application. In many practical cases the performance flexibility inherent in the coupled resonator reflex klystron will more than outweigh this advantage.

It is the purpose of this paper to present the theory underlying the operation of the coupled resonator reflex klystron, as well as its experimental verification. For the sake of completeness, but also in order to emphasize methods of analysis to be used in later sections and not readily found in the literature, a review of reflex klystron theory will precede the exposition of the coupled resonator problem. In both the single and coupled resonator case, performance analysis will be based on a separate and independent study of the electronic and passive circuit admittance developed across the interaction gap and upon the graphical combination of the two in the complex admittance plane. As a by-product of this investigation, a number of driving point properties of two coupled resonant circuits will be developed which may be found of general networks interest. Following the theory of the coupled resonator reflex klystron, an experimental tube of this type will be described and a qualitative as well as quantitative verification of the theory given. Oscillograms will be presented showing the advantages of this device when used as a sweep generator, both in the microwave band and at lower frequencies. Additional applications will be indicated in the hope that others may try them.

2.0 SMALL SIGNAL REFLEX KLYSTRON THEORY

This section will be devoted to a brief review of the small signal reflex klystron theory. Emphasis will be on concepts leading to the equivalent circuit representation and to the graphical admittance-plane analysis. Both of these and particularly the graphical approach will later be used in the investigation of coupled cavity behavior.

As stated before, the operation of the reflex klystron is the result of the interaction between a bunched electron stream and the varying electric field existing inside a resonant cavity. In circuit language, this amounts to an interaction between an active and a passive element. The active one due to the electron stream is termed electronic admittance, and the passive one is the input admittance of the resonator. Derivations for the expression describing the electronic admittance may be readily found in the literature.¹ It will not be repeated here. The re-

sult of these derivations, however, and its physical significance will be discussed at some length.

2.1 Electronic Admittance

Consider an arrangement of four plane and parallel elements consisting of a cathode, two ideal grids and a reflector. Assume these electrodes to be of infinite extent so that all electrons will move in straight paths perpendicular to the planes of the electrodes. Let the current densities encountered be low enough so that space charge effects can be neglected. Both grids are operated at the same dc potential, V_0 , positive with respect to the cathode. As shown in Fig. 1, this might be achieved by connecting them to the secondary winding of an ideal transformer having a 1:1 turns ratio. The reflector is operated at a dc potential, V_R , negative with respect to the cathode. Next, an RF voltage of amplitude V is applied to the transformer and, hence, appears across the grids. Electrons emitted from the cathode are accelerated toward the first grid and arrive at it

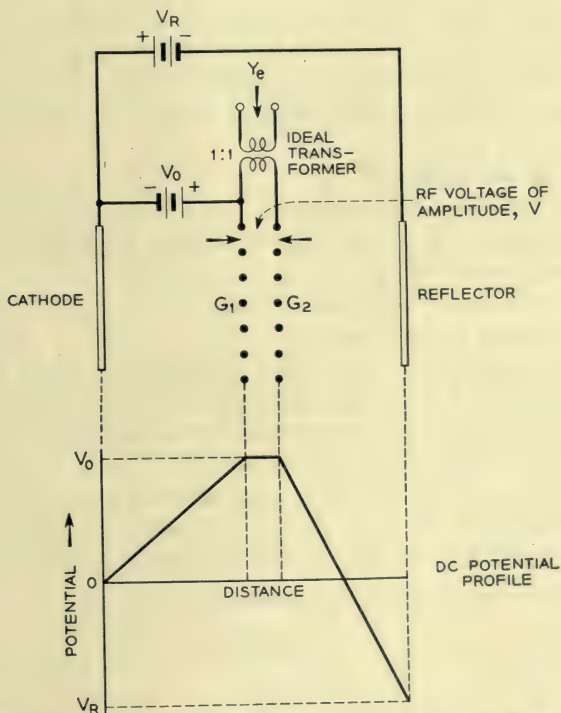


Fig. 1 — Electrode arrangement and dc potential profile giving rise to electronic admittance, Y_e , described by equation (2.1).

with a velocity corresponding to V_0 volts. During their traversal of the interaction gap, i.e., the space bounded by the two grids, this velocity is changed by an amount depending on the instantaneous direction and intensity of the electric field and on the fraction of a cycle spent in traversing the gap. They then enter the retarding dc field of the repeller drift space. Here, they are slowed down, brought to a standstill and returned to the interaction gap for a second transit. During this process, the velocity modulation acquired during the first transit is converted to intensity modulation such that the convection current returning through the gap does so in the form of sharp and well defined pulses with a repetition rate equal to the frequency of the applied RF voltage. This pulsed current may now be harmonically analyzed and its fundamental component evaluated. The ratio of this fundamental component of current to the applied RF voltage is termed electronic admittance, Y_e . Providing the amplitude, V , of the applied RF voltage is small compared to the dc accelerating voltage, V_0 , the value of Y_e is given by:

$$Y_e = \frac{I_0 \beta^2 \theta}{2V_0} \frac{2J_1 \left(\frac{\beta V \theta}{2V_0} \right)}{\left(\frac{\beta V \theta}{2V_0} \right)} e^{j[(\pi/2) - \theta]}, \quad (2.1)$$

where I_0 = dc beam current.

β = beam coupling coefficient.*

θ = round trip transit time in repeller drift space, in radians.

V_0 = dc beam voltage.

V = amplitude of RF gap voltage.

J_1 = Bessel function operator.

Of all these parameters affecting Y_e we shall focus our attention on two, namely V and θ . The other parameters depend on such factors as tube geometry, electron gun perveance etc. and, within the scope of this investigation, will be considered constant.

Referring to equation (2.1), it is seen that the phase angle associated with Y_e is a function of θ only, whereas the amplitude of Y_e depends on both θ and the RF gap voltage, V . Suppose now θ is held constant while

* β , also referred to as modulation coefficient, is given by

$$\beta = \frac{\sin \frac{\theta_g}{2}}{\frac{\theta_g}{2}}$$

where θ_g , the transit angle in the interaction gap, is expressed in radians.

V is increased from zero to some finite value, a process actually occurring in a reflex klystron during the build-up of oscillations. The amplitude of Y_e will then be of the form $2J_1(x)/x$ and will decrease according to the Bessel function plot of Fig. 2.

For an infinitesimal or zero RF gap voltage the electronic admittance, now referred to as "small-signal" electronic admittance, Y_{es} , may be derived from equation (2.1) as,

$$Y_e|_{V=0} \equiv Y_{es} = \frac{I_0 \beta^2}{2V_0} \theta e^{j[(\pi/2)-\theta]}. \quad (2.2)$$

When presented in the complex admittance (i.e., g - b) plane, expression (2.2) assumes the form of a geometric spiral as shown in Fig. 4. Each point on the spiral corresponds to a particular value of θ and, since θ is a function of the repeller voltage only, to a particular value of V_R . We also note that for some values of θ the conductance component of Y_{es} is negative while for others it is positive. Thus, there is the possibility of generation of RF energy for values of Y_{es} adjusted by means of the repeller voltage to fall on the left-hand half of the admittance spiral while energy is absorbed for values of Y_{es} having a positive conductance component. As the RF gap voltage, V , builds up from zero to its final value the magnitude of the electronic admittance shrinks along a radius

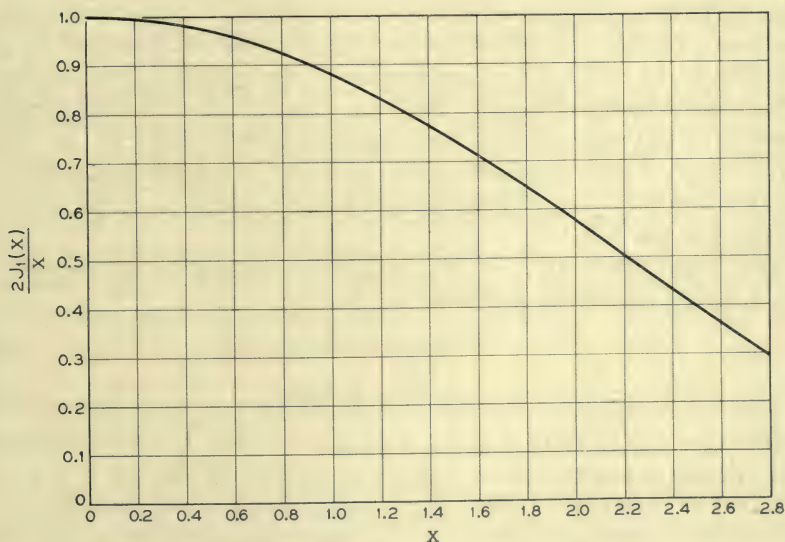


Fig. 2. Bessel function plot showing the relative variation of the amplitude of the electronic admittance (ordinate) as a function of RF gap voltage (abscissa) with repeller drift angle held constant.

vector from Y_{es} to Y_e according to the Bessel function plot of Fig. 2. For $\theta = 2\pi(N + \frac{3}{4})$ radians where $N = 0, 1, 2, 3 \dots$ the electronic admittance becomes a pure negative conductance as may be seen by putting this relation into equations (2.1) or (2.2).

Summarizing, we have seen that the presence of an electron stream bunched in accordance with the arrangement of Fig. 1 gives rise to an admittance appearing across the grids bounding the interaction space. The phase angle of this admittance is a function only of the repeller drift angle, θ , while its magnitude depends on both θ and the RF gap voltage, V . For values of θ in the vicinity of $2\pi(N + \frac{3}{4})$ radians the conductance component is negative, a necessary condition for the production of sustained oscillations.

2.2 Passive Circuit Admittance of Single Resonator

Any resonant cavity may be represented by a simple parallel G - C - L -combination provided the desired resonance is sufficiently far removed from adjacent ones. In some cases such as in cylindrical or waveguide cavities, to name two, it is difficult to ascribe physical significance to the lumped elements appearing in the equivalent circuit representation. This is not so in the case of a conventional reflex klystron cavity. Since the latter always consists of a re-entrant type resonator, most of the electric field is concentrated in the interaction gap, i.e., the narrow region traversed by the outgoing and returning electrons and bounded by two parallel grids, while the major portion of the magnetic flux resides in the outer cylindrical section. Thus, the effective shunt capacitance appearing in the equivalent circuit is associated primarily with the above grid planes, a minor contribution originating in the fringing field close to the re-entrant post and the residual electric field in the outer cylindrical part of the cavity.

The input admittance, Y , of a high Q resonator when represented by C , L , and G connected in parallel is given by,

$$Y = G(1 + j2Q\delta) \quad (2.3)$$

$$= G + j2C\Delta\omega, \quad (2.4)$$

where $Q = \omega_0 C/G$ and $\delta = \Delta\omega/\omega_0 = \Delta f/f_0 = (f - f_0)/f_0$, and G represents all internal resonator losses plus the external load referred to the gap. Plotted in the complex admittance plane, the locus of the admittance vector with varying frequency is a straight line parallel to the imaginary axis and spaced a distance corresponding to G to its right. The variation in susceptance is directly proportional to the frequency deviation

from resonance. Also, the frequency deviation required to bring about a given change in susceptance is inversely proportional to both C and Q .

2.3 Equivalent Circuit of Single Resonator Reflex Klystron

We saw that the presence of a bunched electron stream between two closely spaced grid planes gives rise to an electronic admittance the properties of which were discussed earlier. Suppose we now let these grids become part of a re-entrant cavity so that they form the boundaries of the interaction space. This will make them elements of significance and common to both the electron stream and the passive circuit admittance. As far as the electron stream is concerned the grids become the terminals across which the electronic admittance, Y_e , is developed and in relation to the resonant circuit they constitute the major portion of the effective

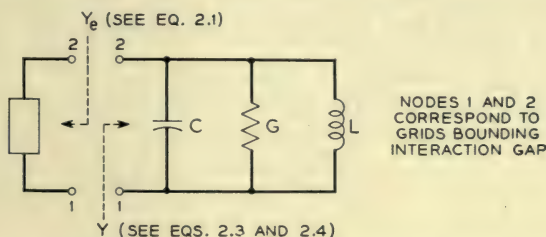


Fig. 3. — Equivalent circuit of single resonator reflex klystron.

shunt capacitance. Based on these remarks, it should be apparent that a single-resonator reflex klystron may be represented by an equivalent circuit consisting of a parallel combination of Y_e , C , G , and L as shown in Fig. 3. Furthermore, the expressions for Y_e and Y as given by equations 2.1 and 2.4, respectively, show that these two quantities are independent of each other. The electronic admittance, Y_e , is a function of the electron optics of the tube, while the passive circuit admittance, Y , is a function purely of cavity parameters, including the tightness of coupling to the external load.

2.4 Performance Analysis Based on Complex Admittance Plane Representation

The condition for oscillation applying to the equivalent reflex klystron circuit requires the total admittance across nodes 1 and 2 of Fig. 3 to equal zero, i.e.,

$$Y_e + Y = 0, \quad 2.5a$$

OR

$$Y_s = -Y.$$

2.5b

Attainment of this condition is preceded by a period during which the RF gap voltage increases from its initial, near-zero value to its final steady state amplitude. Concurrently, the net conductance across nodes 1 and 2 changes from its maximum negative value to zero and the electronic admittance vector shrinks from Y_{es} to Y_s .

This process of build-up of oscillations and the final steady state may be conveniently studied by the graphical representation of Fig. 4. Here, the negative of the passive circuit admittance, Y , has been superimposed on the small-signal electronic-admittance spiral. Since the net conductance across the interaction gap (nodes 1 and 2 in Fig. 3) must be negative in order for oscillations to build up, we see at once that the $(-Y)$ plot, sometimes also referred to as "load line," divides the complex admittance plane into two regions: the one to its left, in which the

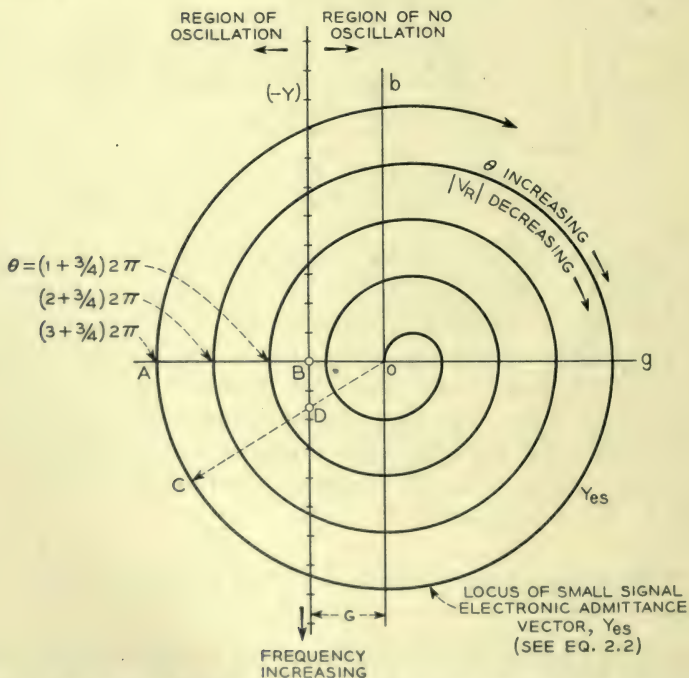


Fig. 4 — Complex admittance plane representation showing the superposition of the negative of the circuit input admittance upon the small signal electronic admittance spiral. Each point on spiral corresponds to a particular value of repeller transit angle, θ , and hence repeller voltage, V_R , and each division on the $(-Y)$ locus to an equal frequency increment.

negative electronic conductance exceeds the circuit conductance and where the build-up of oscillations is possible, and the region to its right where the condition for the build-up of oscillations is not met. Thus, for the load line in the position indicated, the tube will not oscillate in the $N = 0$ mode, but will do so in the $N = 1, 2, 3$ and higher order modes.

Still referring to Fig. 4, suppose θ has been adjusted by means of the repeller voltage to $\theta_1 = (3 + \frac{3}{4})2\pi$ radians, i.e., the center of the $3 + \frac{3}{4}$ mode. The small signal-electronic-admittance vector will then be a pure negative conductance terminating on the spiral at point A and together with the passive circuit conductance yielding a net negative conductance across the grids of value $(\overline{OA} - \overline{OB})$. Oscillations, therefore, will build up until the equilibrium condition, $Y_e = -Y$, has been satisfied. In terms of the admittance plane representation, the electronic-admittance vector will shrink without change in its phase until it terminates on the load line at point B. Since the electronic admittance for the particular value of θ chosen is a pure conductance, oscillations will occur at the resonant frequency of the cavity, f_0 , which is also the frequency corresponding to the intersection of the electronic-admittance vector and the load line.

From equation 2.1 we see that,

$$\frac{\overline{OB}}{\overline{OA}} = \frac{\frac{2J_1 \left(\frac{\beta V_1 \theta_1}{2V_0} \right)}{\left(\frac{\beta V_1 \theta_1}{2V_0} \right)}}{\left[\frac{2J_1 \left(\frac{\beta V \theta}{2V_0} \right)}{\left(\frac{\beta V \theta}{2V_0} \right)} \right]_{V=0}} = \frac{2J_1 \left(\frac{\beta V_1 \theta_1}{2V_0} \right)}{\left(\frac{\beta V_1 \theta_1}{2V_0} \right)}, \quad (2.6a)$$

where V_1 is the steady state RF gap voltage corresponding to drift angle, θ_1 . Entering the graph of Fig. 2 with $\overline{OB}/\overline{OA}$ as the ordinate, we can read off the corresponding value of $(\beta/2V_0)V_1\theta_1$. For a particular tube structure and operating conditions, $\beta/2V_0$ will be a fixed constant so that, in effect, we have determined the value of a quantity proportional to the product of $V_1\theta_1$.

Next, consider the case where θ has been changed from $\theta_1 = (3 + \frac{3}{4})2\pi$ to $\theta_2 = K(3 + \frac{3}{4})2\pi$ radians. The electronic admittance vector for $V = 0$ will now terminate on the Y_{ee} -plot at C. Again a build-up of oscillations will ensue and upon attainment of the steady state condition the vector will have shrunk to the value $\overline{OD}$, with the

frequency of oscillation determined by the location of point D on the load line. By an argument similar to the one used before, we have

$$\frac{\overline{OD}}{\overline{OC}} = \frac{2J_1\left(\frac{\beta V_2 \theta_2}{2V_0}\right)}{\left(\frac{\beta V_2 \theta_2}{2V_0}\right)} \quad (2.6b)$$

and once more, the plot of Fig. 2 will yield the value of

$$\left(\frac{\beta}{2V_0}\right) V_2 \theta_2 \quad \text{or} \quad \left(\frac{\beta}{2V_0}\right) K V_2 \theta_1.$$

It is apparent that $\beta/2V_0$ is a factor common to all these determinations and may be neglected if we are only interested in the mode shape, i.e. the *relative* variation of gap voltage or output power with the repeller drift angle, θ . For modes higher than $N = 2$, the electronic admittance spiral approximates a number of semi-circles with their centers close to the origin. Hence

$$\overline{OA} \approx \overline{OC} \quad \text{and} \quad \frac{\overline{OD}}{\overline{OC}} > \frac{\overline{OB}}{\overline{OA}}.$$

Figure 2, then indicates that

$$V_1 \theta_1 > V_2 \theta_2 \quad \text{or} \quad V_1 \theta_1 > K V_2 \theta_1$$

and since K does not vary greatly from unity $V_1 > V_2$. This is as it should be since we know that a change in repeller voltage from its mid-mode value causes a decrease in power output and, hence, in gap voltage. The latter will decrease to zero when θ has been adjusted to a value such that the Y_{es} vector terminates at the intersections of the electronic admittance spiral and the load line.

Another result which becomes apparent from an inspection of Fig. 4 is this. For the condition of stable oscillation the phase angle of the electronic admittance equals that of the passive circuit except for an additive constant of 180 degrees which, however, may be disregarded in this argument. The frequency of oscillation may be determined from the input-phase-angle vs. frequency plot of the passive circuit by looking up the frequency corresponding to the particular value of phase angle to which the electronic admittance has been adjusted (by means of repeller voltage). Thus, the curve relating repeller-drift-angle to frequency is identical with the plot of input-phase-angle vs. frequency for the passive circuit. Moreover, if repeller voltage is linearly related

to the repeller-drift-angle, a condition which in most practical cases holds over a restricted repeller voltage swing about its midmode value, then the repeller voltage-frequency plot will have the same shape as the input phase angle-frequency curve of the passive circuit, differing from the latter only by a constant multiplying factor. We therefore conclude that the modulation performance of the reflex klystron, at least over the central portion of the mode, may be predicted from an examination of the driving point properties of the passive circuit.

The above method of graphical analysis is quite useful in the case of conventional reflex klystrons, although the same results may be obtained analytically by making use of the equation for electronic admittance in conjunction with that of the input admittance of a single resonator. In the case of coupled resonators the expression for input admittance becomes much more involved, as we shall see later, with the result that the graphical approach outlined above was found by far the quicker and more practical method of solution.

3.0 THEORY OF COUPLED RESONATOR REFLEX KLYSTRON

It has been shown that the performance of a reflex klystron can be analyzed by considering the electronic and passive circuit admittances separately and then combining the two graphically in the complex admittance plane. The same procedure can be adopted in the determination of mode shapes resulting from the interaction of the electronic admittance with any arbitrary circuit admittance which can be realized across the gap. Conversely, we can determine the admittance or impedance function required to produce a particular desired mode shape. In other words: given an admittance function, we can determine the resulting mode shape, and given a desired mode shape, we can predict the required admittance function. As an example, consider a mode having a flat top and vertical sides as shown in Fig. 5(a). This would be the ideal shape for a reflex oscillator to be used as an electronically swept signal source. To achieve this mode shape, we must realize an admittance function across the bunching grids yielding a constant excess of negative electronic admittance over a range of repeller voltages. Such a function is shown in the complex admittance plane along with the plot of Y_{ee} in Fig. 5(b) and the corresponding input impedance in 5(c). It will result in a frequency range of constant RF gap voltage which in turn will give rise to a range of flat power provided it is developed across a constant, frequency-invariant conductance. In order to clarify this rather important consideration, let us write the admittance appearing

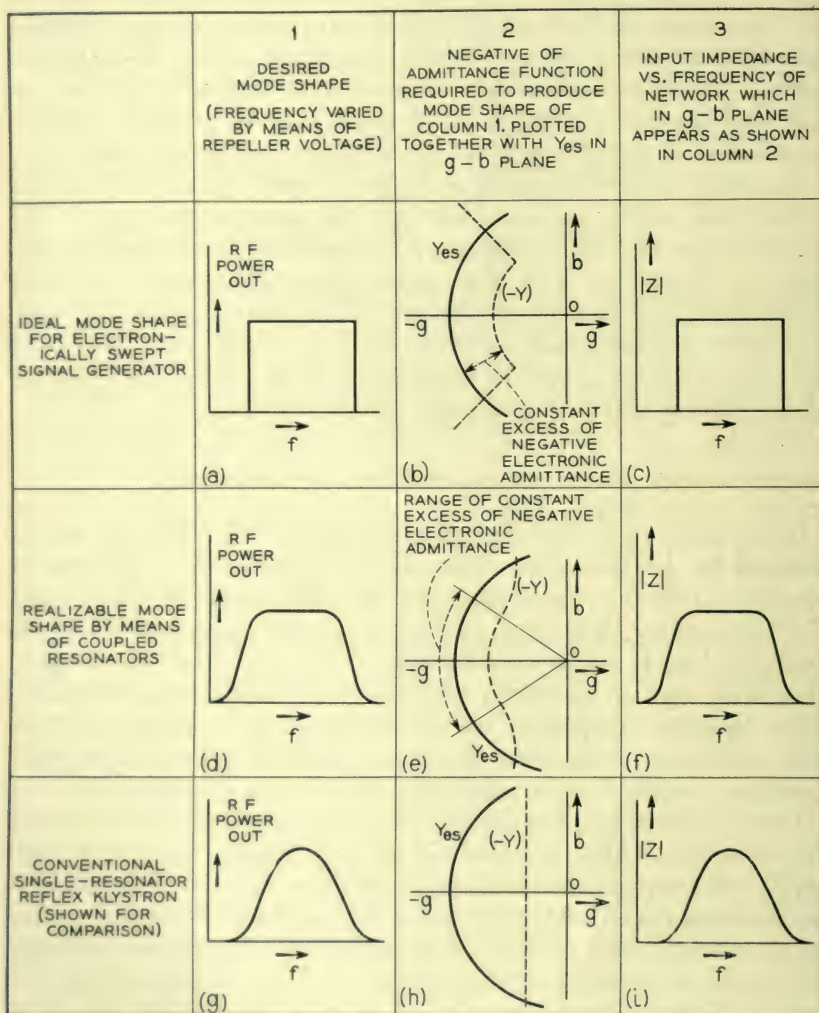


Fig. 5 — Desired mode shapes and passive circuit impedance functions required to produce them.

across the interaction gap as,

$$Y = G_T(f) + jB_T(f),$$

where $G_T(f)$ and $B_T(f)$ denote the total conductance and total susceptance both of which are functions of frequency. Over the region of constant RF gap voltage we require $|Y|$ or

$$\sqrt{G_T^2(f) + B_T^2(f)}$$

to be constant. This, however, means that the power generated (as distinct from the power delivered to the load) is not constant since it is given by,

$$\begin{aligned} \text{Generated Power} &= \frac{1}{2}(\text{RF gap voltage})^2 (\text{Total Conductance}) \\ &= \frac{1}{2}V^2 G_T(f) \end{aligned}$$

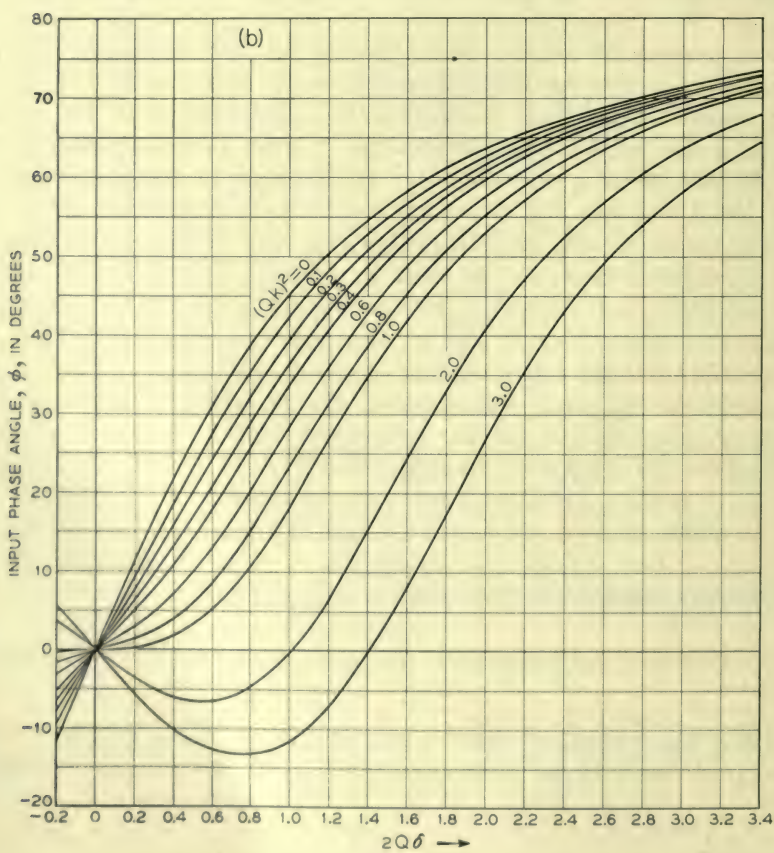
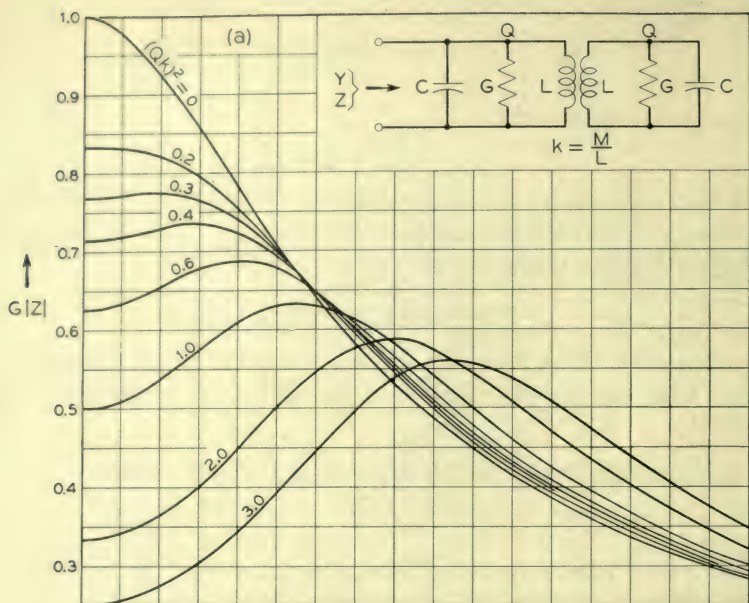
and G_T varies with frequency. Now, the term $G_T(f)$ is the sum of a number of conductances, all but one of which do not change with frequency. The frequency-invariant conductances represent the different power losses in the primary resonator plus the external load referred to the gap while the remaining frequency-sensitive conductance is due to the coupled circuit used in shaping the admittance locus. Hence the useful power output is proportional to V^2 and therefore constant if V^2 is constant, whereas the variation with frequency of the total conductance must be taken into account in evaluating generated power.

The foregoing discussion together with the illustrations of Fig. 5 should make it clear that in order to maintain a constant RF power output level, over a specified frequency range, we must have the electronic admittance interact with a circuit the input impedance of which, when referred to the gap, is also constant over the same frequency range. A circuit having such characteristics can be obtained by the use of coupled resonators as will be shown later.

Suppose the emphasis is on modulation linearity rather than flatness of power output. Attention, then, must be focused on the relation between input phase angle and frequency of the passive circuit. Here, again, we shall see that the application of coupled resonators offers advantages beyond what is possible with a single cavity.

3.1 Driving Point Properties of Two Coupled Resonators Having Equal Q 's

Using the equivalent shunt representation as shown in Fig. 6, the exact expression for the input admittance, Y , of two coupled resonators



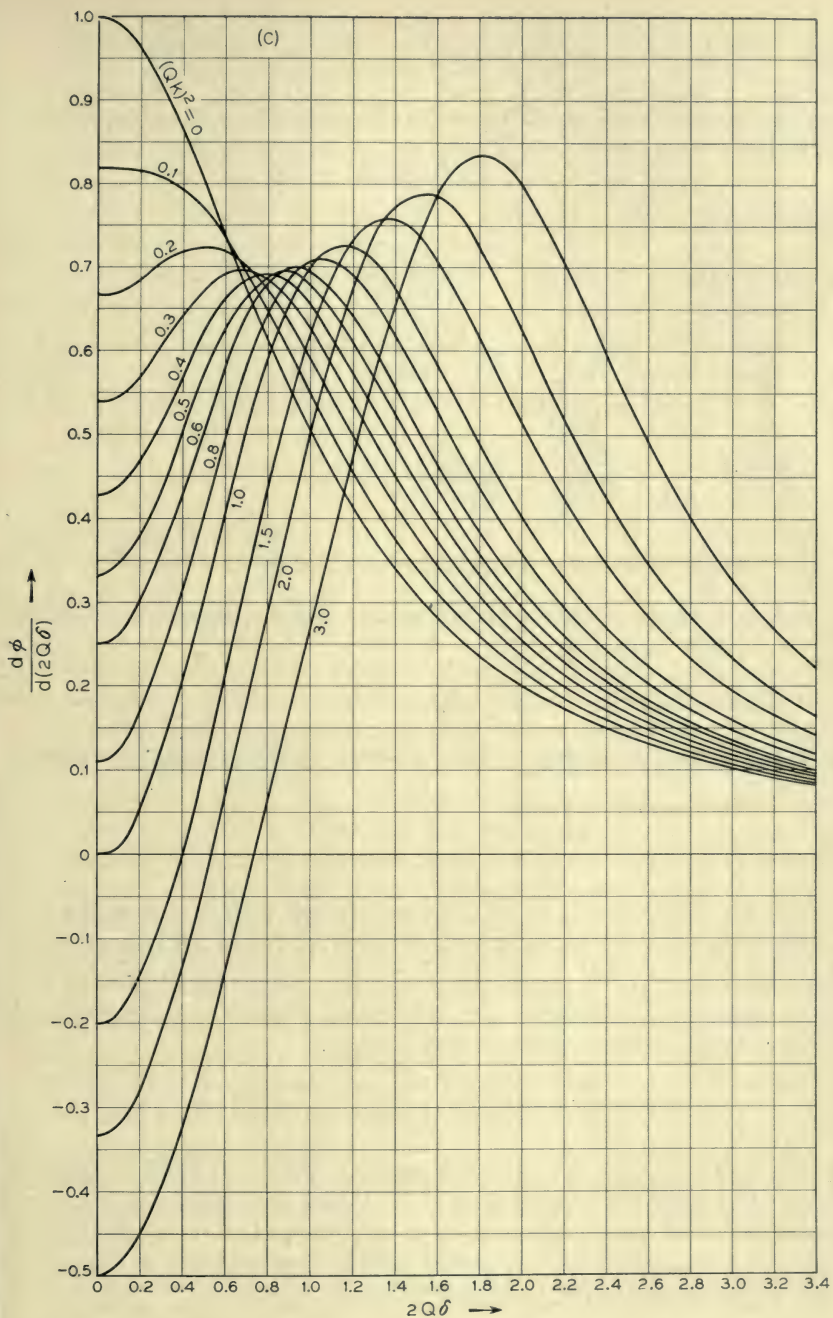


Fig. 6 — Normalized driving point properties of two coupled, synchronously tuned resonators having equal Q 's. (a) Input impedance, $G|Z|$, versus frequency deviation from resonance, $2Q\delta$, for various degrees of coupling, Qk . (b) Input phase angle, ϕ , versus frequency deviation. (c) Rate of change of ϕ with $2Q\delta$.

having equal Q 's and equal resonant frequencies is given by,

$$Y = G(g + jb),$$

where

$$b = \frac{Q(\Omega^2 m - 1)}{\Omega m} \left(1 - \frac{\left(\frac{Qk}{\Omega m}\right)^2}{1 + \left[\frac{Q(\Omega^2 m - 1)}{\Omega m}\right]^2} \right), \quad (3.1)$$

$$g = 1 + \frac{\left(\frac{Qk}{\Omega m}\right)^2}{1 + \left[\frac{Q(\Omega^2 m - 1)}{\Omega m}\right]^2}.$$

G = shunt conductance of primary and secondary resonators.

g, b = total input conductance and susceptance respectively normalized with respect to G ,

Ω = normalized frequency = f/f_0 .

f_0 = resonant frequency of both primary and secondary resonators.

k = coefficient of coupling.

$m = 1 - k^2$.

This expression for input admittance, though accurate, is rather unwieldy because of the many variables involved. A few obvious approximations, however, will change equation (3.1) into a much simpler and more meaningful expression, yet sufficiently accurate for the range of Q 's and bandwidths of interest here. Let,

$$\Omega = \frac{f}{f_0} = \frac{f_0 + \Delta f}{f_0} = 1 + \frac{\Delta f}{f_0} = 1 + \delta,$$

where δ denotes the normalized frequency deviation from resonance, $\Delta f/f_0$. If we further assume that $k^2 \ll 1$ and the range of Q 's is such that $(Qk)^2$ may vary from zero to about five and that the maximum value of δ is small enough so that its higher powers may be neglected, then equation (3.1) may be simplified to,

$$\frac{Y}{G} = \left[1 + \frac{(Qk)^2}{1 + (2Q\delta)^2} \right] + j2Q\delta \left[1 - \frac{(Qk)^2}{1 + (2Q\delta)^2} \right]. \quad (3.2)$$

The above equation* essentially contains three variables:

- (a) The dependent variable, Y/G , i.e. input admittance normalized with respect to the shunt conductance,
- (b) the independent variable, $(2Q\delta)$, which is a factor proportional to the frequency deviation from resonance, and
- (c) parameter, (Qk) , a measure of the tightness of coupling between the two cavities.

Compared with the input admittance for a single resonator which was given earlier, [equation (2.3),] as, $Y/G = 1 + j2Q\delta$, it is seen that the conductance component has been changed by a factor

$$\left[1 + \frac{(Qk)^2}{1 + (2Q\delta)^2} \right]$$

and the susceptance by

$$\left[1 - \frac{(Qk)^2}{1 + (2Q\delta)^2} \right].$$

Also, by setting $k = 0$, i.e., completely decoupling the secondary resonator, equation (3.2) reduces to equation (2.3) as, indeed, it should.

The information contained in equation (3.2) may be presented graphically in a number of ways. We can plot the magnitude of the normalized input impedance, $G|Z|$, as a function of $2Q\delta$ with $(Qk)^2$ as parameter as shown in Fig. 6(a).† Or we can plot the input admittance given by equation (3.2) directly in the g - b plane as in Fig. 7. Finally, we may show the variation of input phase angle with normalized frequency for different degrees of coupling as in Fig. 6(b). Each of these graphical representations has an important bearing on the performance of the coupled resonator klystron. Thus, the curves of Fig. 6(a) will have the same general shape as the RF power output vs. frequency plot of the reflex oscillator, the family of curves of Fig. 7, when superimposed on the electronic admittance spiral, can be used for a detailed analysis of mode shapes, and Fig. 6(b) determines the modulation characteristics obtainable with coupled resonators.

Having briefly touched upon the significance of the families of curves given in Figs. 6(a), 7, and 6(b), we can now proceed to discuss them in greater detail and to establish further valuable results.

* To check the accuracy of this equation, the curve for $(Qk)^2 = 1$ in Fig. 6(a) was replotted using the full expression given by equation 3.1 and assuming $Q = 100$. The agreement was found to be essentially perfect as far as its use in this investigation is concerned.

† For a similar presentation of the *transfer* characteristics of coupled tuned circuits see reference (6).

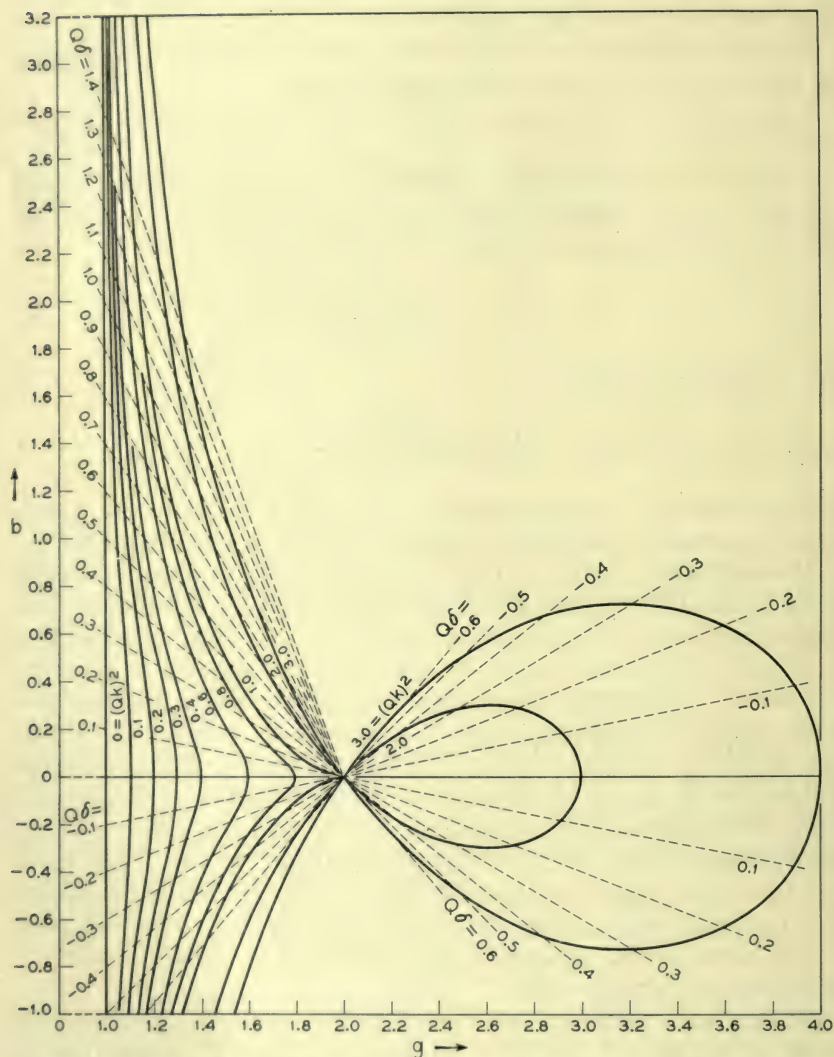


Fig. 7 — Input admittance of circuit shown in Fig. 6 plotted in g - b plane. Solid lines are loci of input admittance vectors, dashed lines connect points of equal frequencies.

3.1.1 *Variation of Input Impedance with Frequency, Fig. 6(a).* Equation 3.2 may be written as

$$\frac{Y}{G} = g + jb, \quad (3.3)$$

where

$$g = \left[1 + \frac{(Qk)^2}{1 + (2Q\delta)^2} \right] \quad (3.4)$$

and

$$b = 2Q\delta \left[1 - \frac{(Qk)^2}{1 + (2Q\delta)^2} \right]. \quad (3.5)$$

Hence

$$\left| \frac{Y}{G} \right| = \sqrt{g^2 + b^2}$$

and

$$G|Z| = \frac{1}{\sqrt{g^2 + b^2}}. \quad (3.6)$$

It is seen that g^2 and b^2 involve even powers of $(2Q\delta)$ only, so that $G|Z|$ is an even function, i.e., symmetrical about the vertical axis. For this reason, one-half the normalized impedance plot only has been given in Fig. 6(a). Inspection of this figure reveals that the effect of coupling to a secondary resonator is to broaden the frequency range over which a high impedance level can be maintained across the interaction gap. Comparing the variation of input impedance with frequency for $(Qk)^2 = 0$ (i.e., single cavity) with that for $(Qk)^2 = 0.3$, we see that for a frequency deviation of $2Q\delta = \pm 0.6$ the former shows a drop of 14 per cent while the latter only varies by ± 0.58 per cent. In terms of two coupled resonators having $Q = 100$ and operating at 4000 mc this means a variation in impedance of only ± 0.58 per cent or ± 0.052 db over a frequency range of 24 mc.

Another result clearly brought out by the family of curves of Fig. 6(a) is that the process of broadbanding by coupling to a second resonator results in a reduction in absolute impedance level. This reduction in impedance level at midband is related to the tightness of coupling, $(Qk)^2$, by the expression,

$$[G|Z|]_{\delta=0} = \frac{1}{1 + (Qk)^2}, \quad (3.7)$$

which may be readily derived from eq. 3.6. Note that for $(Qk)^2 = 1$, the impedance level at midband has dropped to half the value obtained with a single resonator or $(Qk)^2 = 0$. The degree of coupling corresponding to $(Qk)^2 = 1$ is noteworthy for reasons other than the one just mentioned. They will be discussed in the following section.

3.1.2 Input Admittance Plot in g - b Plane, Fig. 7. The graphical representation of input admittance in the complex admittance plane is of particular usefulness in the analysis of the coupled resonator reflex klystron. For the moment, however, we shall restrict the discussion to a consideration of the passive circuit only.

Each solid line in Fig. 7 is the locus of the admittance vector for a particular tightness of coupling. For $(Qk)^2 = 0$ the locus is a straight line parallel to the susceptance axis as described earlier for the case of a single resonator. For $(Qk)^2 = 0.3$ the shape of the locus approximates a circle with its center at the origin; this, of course, being true only over a restricted frequency range. As the coupling to the secondary resonator is progressively tightened, the locus is seen to bulge out in the direction of increasing conductance until, for $(Qk)^2 = 1$, it forms a cusp. This condition will henceforth be referred to as "critical coupling."* Coupling even tighter causes the formation of loops of increasing size.

For the overcoupled case, $(Qk)^2 > 1$, the admittance-vector locus crosses the conductance axis three times, with the first and third crossings coincident and independent of $(Qk)^2$ and the second crossover a function of $(Qk)^2$. The location of these intersections with the g -axis may be determined by equating the susceptance to zero, i.e., from equation (3.5),

$$2Q\delta \left[1 - \frac{(Qk)^2}{1 + (2Q\delta)^2} \right] = 0.$$

Hence the crossing to the extreme right occurs for $2Q\delta = 0$ while the first and third intersections correspond to

$$2Q\delta = \pm \sqrt{(Qk)^2 - 1}.$$

To determine the size of the loop we substitute $2Q\delta = 0$ into the expression for the conductance, i.e., equation (3.4) and obtain,

$$g|_{2Q\delta=0} = 1 + (Qk)^2,$$

* It should be noted that the term "critical coupling" as applied to the transfer characteristics of coupled tuned circuits, though also occurring for $(Qk)^2 = 1$, assumes a different significance in that it describes the condition of maximum flatness in response and optimum phase linearity. (See reference 6.) In the case of the coupled resonator reflex klystron, "critical coupling" forms the transition between stable performance and load hysteresis as will be shown later.

showing that the size of the loop is a sensitive function of the degree of coupling. The conductance value for the first and third crossover points is obtained by substituting

$$2Q\delta = \pm\sqrt{(Qk)^2 - 1}$$

into equation (3.4) and results in $g = 2$, i.e., a value of g independent of the coupling coefficient.

The dashed lines shown in Fig. 7 connect points of equal frequencies. It is of interest to note that these loci are straight lines crossing the conductance axis at $g = 2$. To prove this, eliminate $(Qk)^2$ between equations (3.4) and (3.5). This yields,

$$g - 1 = 1 - \frac{b}{2Q\delta},$$

whence

$$b = (-2Q\delta)(g - 2). \quad (3.8)$$

For a particular and constant value of $2Q\delta$, equation (3.8) describes a straight line of slope equal to $(-2Q\delta)$ intersecting the g -axis at $g = 2$.

3.1.3 Variation of Input Phase Angle with Frequency, Fig. 6(b). The last driving point property of interest to this study is the dependence upon frequency of the input phase angle, ϕ . Referring to equation (3.2), this quantity is obtained as,

$$\phi = \tan^{-1} \left[2Q\delta \frac{1 + (2Q\delta)^2 - (Qk)^2}{1 + (2Q\delta)^2 + (Qk)^2} \right]. \quad (3.9)$$

The graphical representation of this function is given in Fig. 6(b). It shows the gradual transition from a simple S-shaped curve for $(Qk)^2 = 0$, having its only point of inflection at the origin, to the type of curve corresponding to $(Qk)^2 > 1$ which intersects the frequency axis three times. The special case of $(Qk)^2 = 1$, considered earlier and found to result in the formation of a cusp in the complex admittance plane, now gives rise to a plot of input phase which is tangent to the horizontal axis at the origin.

To investigate the condition for greatest linearity between phase angle and frequency, which, when applied to the coupled resonator reflex klystron, would be the condition for optimum modulation linearity, one could simply apply a straight edge to the curves of Fig. 6(b) and pick the best value of $(Qk)^2$ in this manner. A much more sensitive criterion of linearity, however, is the variation with frequency of the slope of these curves. An analytical expression for this slope has been

derived from equation (3.9) as,

$$\frac{d\phi}{d(2Q\delta)} = \frac{[1 - (Qk)^4] + (2Q\delta)^2[2 + 4(Qk)^2] + (2Q\delta)^4}{[1 + (Qk)^2]^2 + (2Q\delta)^2[3 + (Qk)^4] + (2Q\delta)^4[3 - 2(Qk)^2] + (2Q\delta)^6} \quad (3.10)$$

The above expression, involving even powers of $2Q\delta$ only, results in a family of curves symmetrical about the vertical axis, the positive half of which is shown in Fig. 6(c). From it the value of $(Qk)^2$ for greatest linearity or most constant slope is seen to lie somewhere between 0.1 and 0.2. This is further borne out by Fig. 8, in which this region has been more fully explored. Fig. 8 constitutes a plot quite similar to that

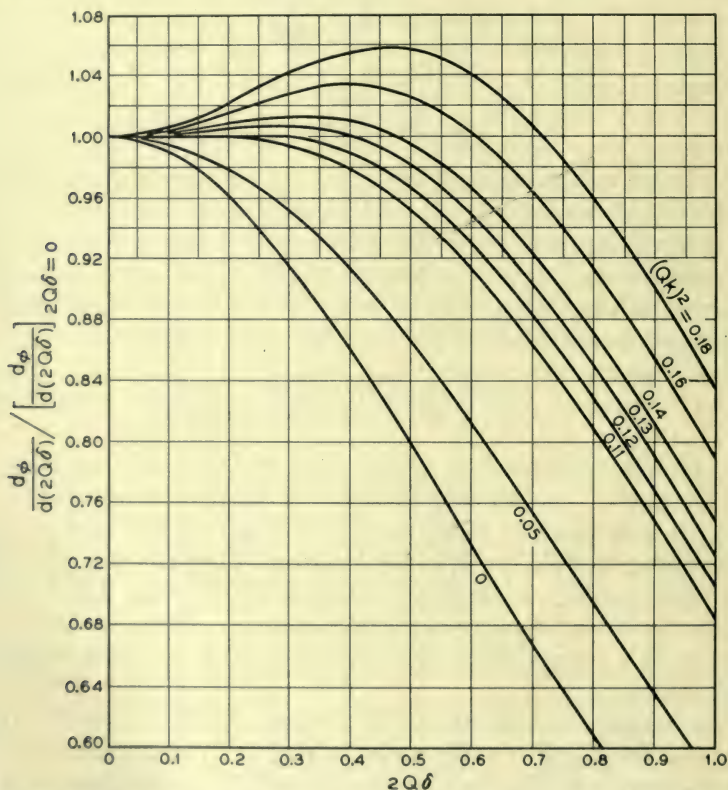


Fig. 8 — Rate of change of input phase angle with frequency normalized with respect to its center value for the circuit shown in Fig. 6. This quantity, the constancy of which is a measure of phase angle versus frequency linearity, is seen to stay absolutely constant over a frequency range corresponding to $2Q\delta = \pm 0.3$ for $(Qk)^2 = 0.12$ while the single resonator case, i.e. $(Qk)^2 = 0$, shows a change of 8% over the same frequency range.

of Fig. 6(c) except that the ordinate now represents the instantaneous slope normalized with respect to its midmode value. The curves shown are for values of $(Qk)^2$ ranging from zero to $(Qk)^2 = 0.18$. Let us, for example, examine the curve corresponding to $(Qk)^2 = 0.12$; the slope is seen to stay absolutely constant up to a frequency deviation of $2Q\delta = \pm 0.3$, while the plot for the single resonator, i.e., $(Qk)^2 = 0$, included in this figure for comparison, changes by 8.4 per cent over the same frequency range.

Putting this in another way, suppose the maximum allowable deviation in slope from its mid-band value is one per cent. Fig. 8 then indicates that the permissible frequency deviation for coupled resonators having a value of $(Qk)^2 = 0.135$ is given by $2Q\delta = \pm 0.5$, whereas it must be restricted to one-fifth this value, i.e., $2Q\delta = \pm 0.1$, for the single resonator case. In terms of a midband frequency of 4000 mc and $Q = 100$, the permissible frequency excursions would be ± 10 mc and ± 2 mc, respectively.

It is to be noted that the value of coupling coefficient resulting in greatest modulation linearity is considerably smaller than the value of coupling coefficient found to yield a constant impedance level.

3.2 Mode Shapes Resulting from the Interaction Between Electronic Admittance and Input Admittance of Two Coupled Resonators of Equal Q 's

Having determined all relevant properties of the passive circuit as they appear across the grids of the bunching gap, we may now proceed to investigate the results arising from their interaction with the electronic admittance. The approach to be adopted is essentially the same as the one outlined earlier for the case of the conventional single-cavity reflex klystron. It involves a graphical superposition of the negative of the passive circuit admittance upon the small-signal-electronic-admittance spiral in the g - b plane, such as shown in Fig. 9(a). The location of the load lines with respect to the spiral has been chosen, somewhat arbitrarily, such that the ratio between the length of the Y_{ee} -vector corresponding to $\theta = (2 + \frac{3}{4})$ cycles and that of the input admittance vector for $(Qk)^2 = 0$ and $\delta = 0$ equals two. Load lines have been drawn for five values of $(Qk)^2$ ranging from zero to unity.

The determination of mode shapes from Fig. 9(a) proceeds as illustrated in Table I. Taking the repeller drift angle as the independent variable we can obtain corresponding values of generated power (in arbitrary units) and frequency (in terms of $Q\delta$) by going through the steps indicated.

DETERMINATION OF THEORETICAL MODE SHAPES FOR $3\frac{3}{4}$ REPELLER MODE AND SECONDARY Q EQUAL TO PRIMARY Q

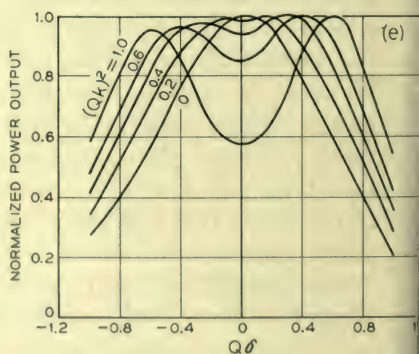
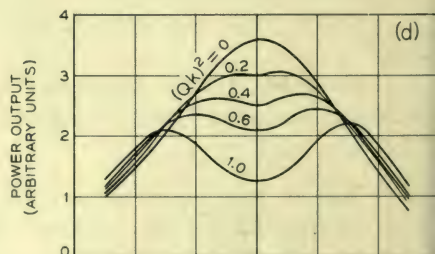
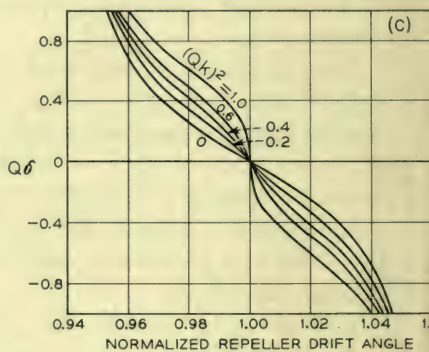
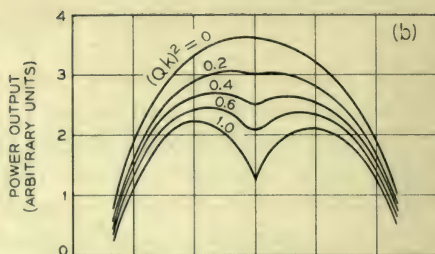
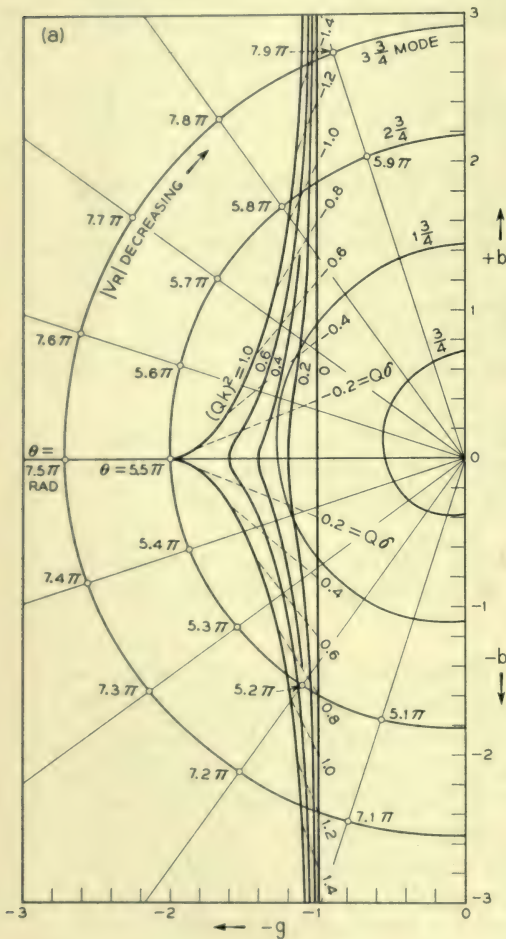


Fig. 9 — Graphical determination of mode shapes for a coupled resonator reflex klystron having the secondary cavity Q equal to primary Q . (a) Load lines for various degrees of coupling superimposed on small signal electronic admittance spiral. (b) Power output versus repeller drift angle. (c) Frequency versus repeller drift angle. (d) Power output versus frequency. (e) Power output, normalized with respect to its maximum value within the mode, shown as a function of frequency.

TABLE I — DETERMINATION OF MODE SHAPES FROM COMPLEX ADMITTANCE PLANE PLOT OF FIGURE 9(a)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$K\theta_0$, Repeller Drift Angle (θ_0 = Midmode Value)	K , Ratio of Drift Angle to its Midmode Value	Small Signal Elec- tronic Admit- tance	Passive Circuit Admit- tance	$\frac{2J_1(KV\theta_0)}{(KV\theta_0)}$ i.e. (4)/(3)	($KV\theta_0$) from Bessel- function plot of Fig. 2	$V\theta_0$, (6)/(2)	($V\theta_0$) ² , Quantity Proportional to Power Output	$Q\delta$
$(Qk)^2 = 0.4$, $2\frac{3}{4}$ -mode, i.e., $\theta_0 = (2 + \frac{3}{4})2\pi$ radians = 5.5π radians								
		Measured off Fig. 9(a) in Arbitrary Units of Length						
$\theta_0 = 5.50\pi$	1.000	10.00	7.00	0.700	1.65	1.645	2.71	0.00
5.20 π	0.947	9.45	9.45	1.000	0.00	0.000	0.00	+0.85
5.25 π	0.955	9.54	8.10	0.849	1.13	1.182	1.40	+0.67
5.30 π	0.967	9.63	7.35	0.764	1.44	1.493	2.23	+0.53
5.35 π	0.974	9.72	6.95	0.715	1.60	1.645	2.70	+0.42
5.40 π	0.983	9.82	6.80	0.693	1.67	1.700	2.89	+0.30
5.45 π	0.992	9.91	6.80	0.686	1.69	1.703	2.90	+0.17
5.55 π	1.010	10.10	6.80	0.673	1.73	1.713	2.94	-0.17
5.60 π	1.020	10.20	6.80	0.667	1.75	1.715	2.95	-0.30
5.65 π	1.029	10.29	6.95	0.676	1.72	1.673	2.80	-0.42
5.70 π	1.038	10.38	7.35	0.709	1.62	1.560	2.44	-0.53
5.75 π	1.046	10.46	8.10	0.775	1.40	1.340	1.80	-0.67
5.80 π	1.055	10.55	9.45	0.895	0.95	0.900	0.81	-0.85

The results of this analysis are shown in Fig. 9. This illustration, in addition to giving detailed performance characteristics for the $3 + \frac{3}{4}$ mode, also indicates clearly the wealth of information which may be obtained from the complex admittance plane representation. Although the curves shown are self-explanatory, a few comments regarding their significance would seem to be in order. It is seen, for instance, that a coupling coefficient so adjusted that $(Qk)^2$ lies between 0.2 and 0.4 will produce a frequency range of essentially constant power. In particular, if we pick the curve for $(Qk)^2 = 0.4$ from the family of curves of Fig. 9(d), it will be seen that the variation in power over a bandwidth of $Q\delta = \pm 0.4$ is ± 2 per cent, while the corresponding value for the single resonator case, i.e., $(Qk)^2 = 0$, is minus 23 per cent. As the value of $(Qk)^2$ is increased beyond 0.4, the depression in the center of the mode becomes increasingly pronounced until it turns into a cusp for $(Qk)^2 = 1$. Mode shapes for $(Qk)^2 > 1$, though of no direct interest to this investigation, are indicated in Fig. 10 since they may be encountered in practice in cases of excessively tight coupling and could then be recognized as such. If the mode is traversed in the direction of increasing $|V_R|$ (or increasing frequency), such that the intersection of the electronic

admittance vector with the load line moves down along the latter as indicated by the arrows in Fig. 10(a), power changes smoothly until the Y_e -vector becomes tangent with the loop at 4. At this point there occurs a discontinuous jump in both power and frequency caused by the sudden shortening of the Y_e -vector from 4 to 7. From here on, power and frequency again become single valued functions of V_R . The mode shape corresponding to this uni-directional sweep is shown in Fig. 10(b). If the mode is traversed in the opposite sense we again encounter this discontinuity although it will now occur on the opposite side of the loop and, therefore, at a different repeller voltage and frequency. Fig.

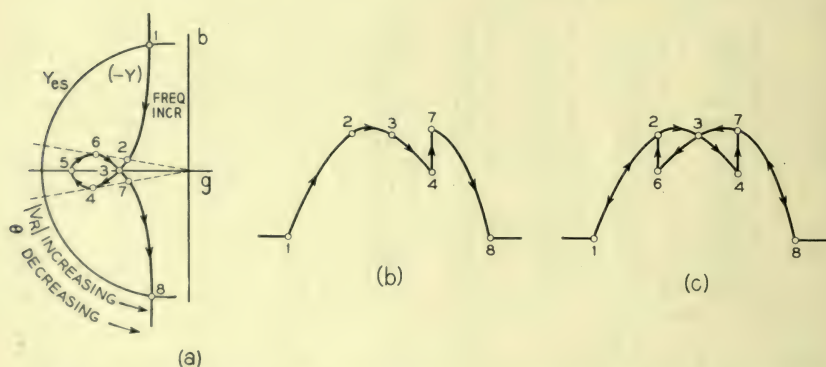


Fig. 10 — Production of load hysteresis by overcoupled cavities. (a) Load line for overcoupled cavities. (b) Oscillographic mode representation for unidirectional (sawtooth) repeller sweep. (c) Oscillographic mode representation for sinusoidal repeller sweep.

10(c) shows the load hysteresis effect as one might expect to observe it on the oscilloscope screen with a sinusoidal sweep applied to the repeller.

3.3 Driving Point Properties of Two Coupled Resonators Having Unequal Q 's

The presentation of the theory of the coupled resonator reflex klystron will now be concluded by a discussion of the more general and, as we shall see, more useful case of two coupled cavities of unequal Q 's. Specifically, we are considering the equivalent circuit shown in Fig. 11. By making approximations similar to those which led to equation 3.2, the input admittance for the case of unequal Q 's may be derived as,^{2, 4, 5}

$$\frac{Y}{G} = \left[1 + \frac{k^2 Q Q_s}{1 + (2Q_s \delta)^2} \right] + j2Q\delta \left[1 - \frac{(Q_s k)^2}{1 + (2Q_s \delta)^2} \right], \quad (3.11)$$

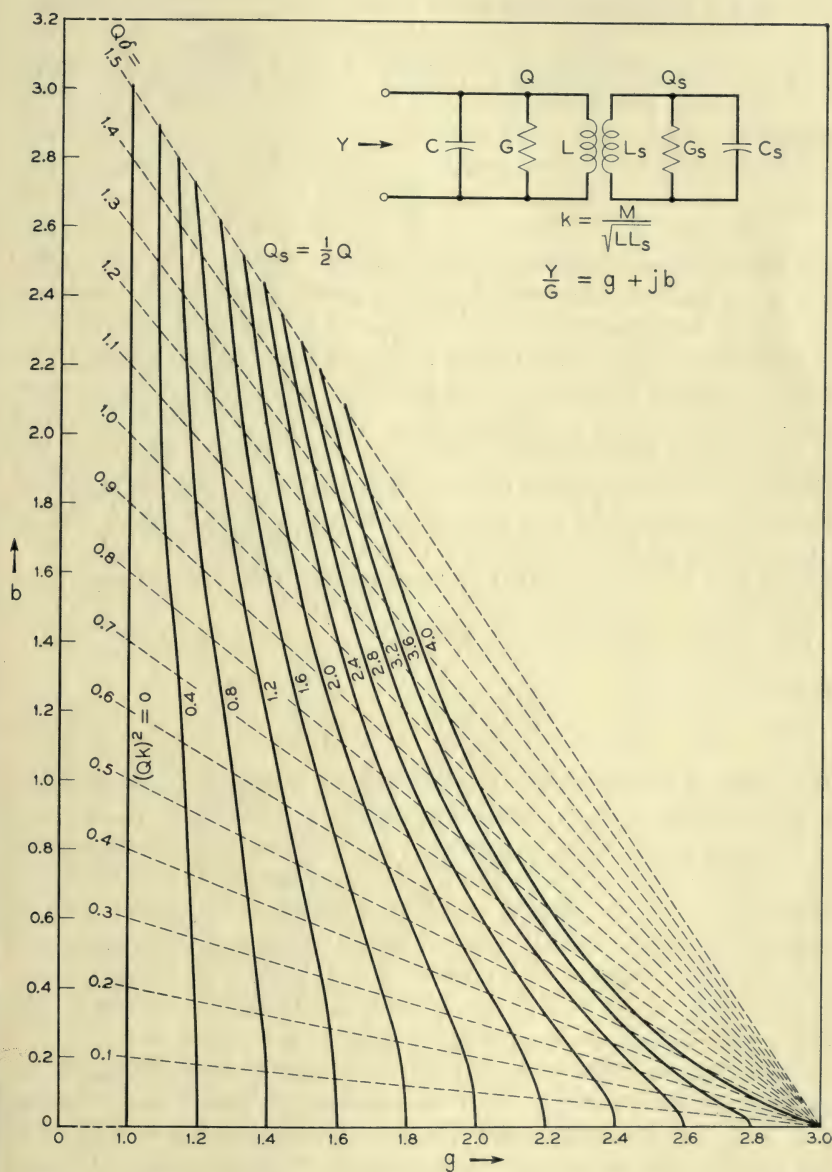


Fig. 11 — Driving point properties of two coupled, synchronously tuned resonators for the case of the secondary Q equal to half the primary Q .

or in more convenient form as,

$$\frac{Y}{G} = \left[1 + \frac{p(Qk)^2}{1 + p^2(2Q\delta)^2} \right] + j(2Q\delta) \left[1 - \frac{p^2(Qk)^2}{1 + p^2(2Q\delta)^2} \right], \quad (3.12)$$

where

$$p = \frac{Q_*}{Q}.$$

The above expression contains four variables, namely Y/G , (Qk) , $(2Q\delta)$ and p , so that the complex admittance plane representation will now have to be restricted to particular values of p . One such plot, for $p = \frac{1}{2}$, appears in Fig. 11. It is similar to that of Fig. 7 except that only the positive half has been shown since equation (3.12) is symmetrical about the conductance axis; in addition, the parameter, $(Qk)^2$, has been carried to the point of critical coupling only. i.e., the formation of a cusp, and not beyond. The frequency contours are again seen to be straight lines crossing the horizontal axis at a value of conductance equal to the input conductance for the condition of critical coupling and $2Q\delta = 0$.

Equation (3.12) shows that the susceptance term will be zero for

$$(a) \quad 2Q\delta = 0,$$

or for

$$(b) \quad p^2(Qk)^2 = 1 + p^2(2Q\delta)^2. \quad (3.13)$$

The value of conductance corresponding to condition (a) is given by $g = 1 + p(Qk)^2$ and the value corresponding to condition (b) by $g = 1 + 1/p$. It is interesting to note that this latter value which determines the point of intersection of the frequency contours, as well as of the admittance plot for critical coupling, with the conductance axis, is independent of the actual values of Q and the degree of coupling and only dependent upon p , the ratio of Q 's. Thus in Fig. 11, which constitutes a plot for $p = 0.5$, the value of conductance at which all the above named contours meet is given by $g = 1 + 1/0.5 = 3$.

From what has been said before we know that at critical coupling the admittance locus forms a cusp intersecting the conductance axis at $g = 1 + 1/p$ at the frequency, $2Q\delta = 0$. Substituting this value of $2Q\delta$ into the conductance term of equation 3.12 and equating to $1 + 1/p$ yields

$$1 + p(Qk)^2 = 1 + \frac{1}{p}.$$

Hence

$$pQk = Q_s k = 1 \text{ (for critical coupling).} \quad (3.14)$$

The effect of reducing the secondary resonator Q is to broaden the frequency range over which a high input impedance or low admittance may be maintained. This is clearly illustrated in Fig. 12 where curves having the same value of conductance at resonance have been selected from Figs. 7 and 11 and superimposed to facilitate comparison.

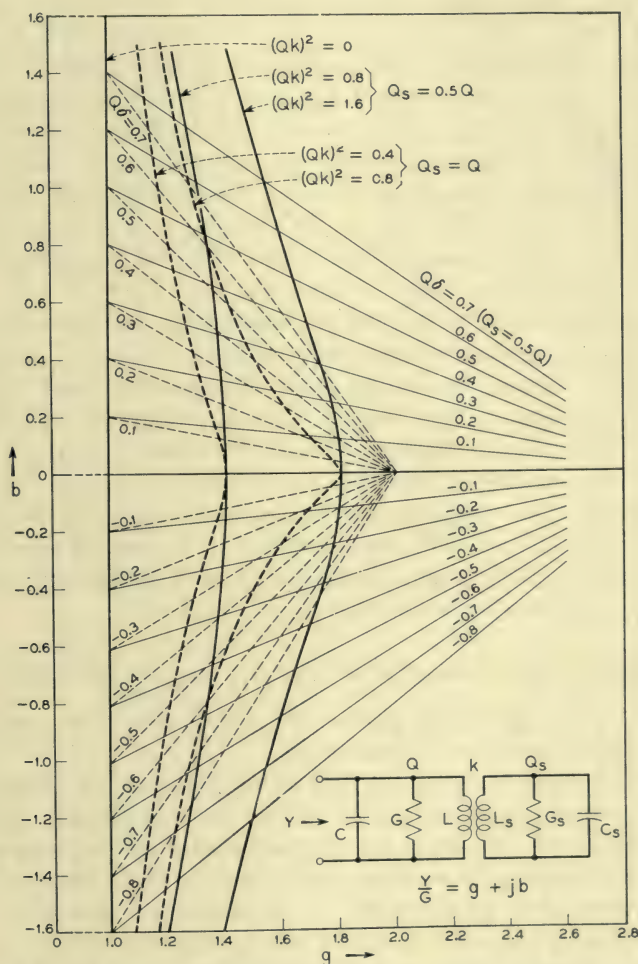


Fig. 12 — Comparison of admittance loci for coupled resonators of equal Q 's with the case in which the secondary Q equals half the primary Q .

3.4 Mode Shapes Obtainable with Two Coupled Resonators of Unequal Q 's

The complex admittance plane representation for the case of the secondary Q equal to half the primary Q is shown in Fig. 13(a) and the resulting mode shapes for the $3 + \frac{3}{4}$ repeller mode in Figs. 13(b), (c), (d), and (e). Again we notice that the coupling coefficient required for best modulation linearity is considerably smaller than that which results in a flat topped power curve. From Fig. 13(c) the most linear frequency-repeller voltage curve is associated with $(Qk)^2 \cong 0.8$ while from Fig. 13(d) a flat power mode may be obtained with a value of $(Qk)^2$ somewhat less than 1.6. In cases where neither modulation linearity nor constant power output are of importance but where the application requires a wide electronic tuning range, Fig. 13(e) shows the advantage to be gained from coupled cavities. Here, power output has been normalized with respect to its peak value within the particular mode under consideration and plotted against $Q\delta$. The ratio of half-power bandwidths for the curve corresponding to $(Qk)^2 = 2.4$ to the single resonator case, i.e., $(Qk)^2 = 0$, is seen to equal 1.73.

A phenomenon which may be encountered in the operation of the coupled resonator reflex klystron is illustrated by the $(Qk)^2 = 4.0$ curve of Fig. 13(b). Here we are dealing with a split mode in which the power, though everywhere a single valued function of repeller voltage, drops to zero over a range of repeller voltages centered about the middle of the mode. The reason for this behavior may be readily understood from an inspection of the complex admittance plane representation of Fig. 13(a). It is caused by the Y_{ss} -locus for the $3 + \frac{3}{4}$ repeller mode crossing the appropriate load line and thereby resulting in a frequency band over which the condition for oscillation cannot be met.

4.0 AN EXPERIMENTAL COUPLED-RESONATOR REFLEX KLYSTRON

The reduction to practice of the theoretical results obtained in the above study raises these requirements:

- (1) An arrangement must be found which allows the coupling between primary and secondary cavities as well as between primary cavity and waveguide output line to be varied continuously and independently.
- (2) Either primary or secondary resonators (or both) must be tunable to allow frequency adjustments for synchronous operation.
- (3) Secondary resonator Q should be continuously variable.
- (4) The secondary resonator should be detachable for independent determination of Q .

DETERMINATION OF THEORETICAL MODE SHAPES
FOR $3\frac{3}{4}$ REPELLER MODE AND SECONDARY Q
EQUAL TO PRIMARY Q

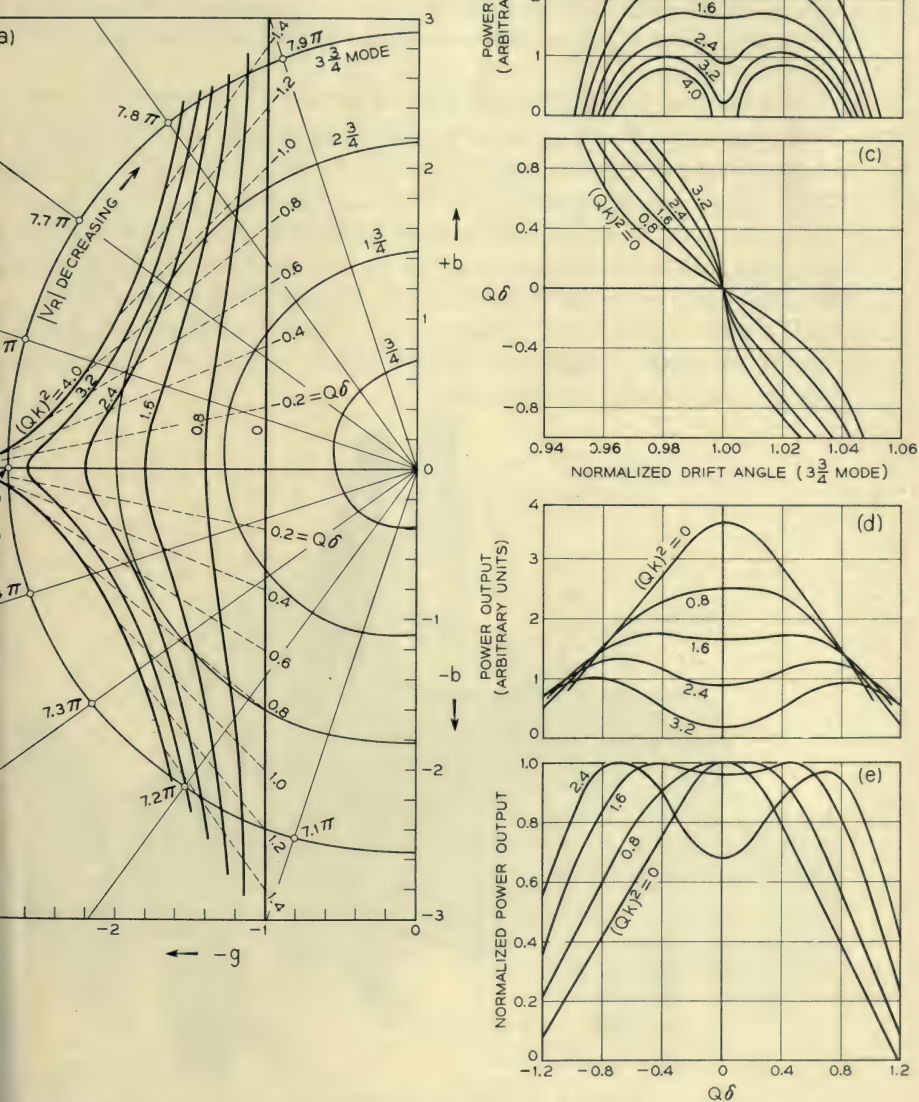


Fig. 13 — Graphical determination of mode shapes for a coupled resonator reflex klystron having the secondary cavity Q equal to half the primary Q . The effect of lowering the secondary Q may be studied by comparing the above curves with the corresponding ones shown in Fig. 9.

(5) Frequency at which these experiments are to be performed should be in a microwave band where good waveguide components and measurement techniques are available.

These considerations have led to the adoption of an external cavity type reflex klystron, the Sylvania 6BL6, as a vehicle for the experimental studies to be described, and operation in the 3700–4200 mc band.

4.1 Constructional Features of Experimental Tube and Circuit

The 6BL6 is one of a group of Sylvania low-voltage reflex klystrons¹⁰ designed for use with external cavity resonators. Electrical connection to the interaction-gap grids is made through gold plated contact rings, formed from the disc seals which pass through the glass. The top ring is slightly smaller in diameter than the bottom ring, thereby permitting the insertion of the tube without disturbance to the associated cavity. Fig. 14 shows the external appearance of this tube and also contains a view of the major components of the passive circuit with cutouts to indicate the internal construction.

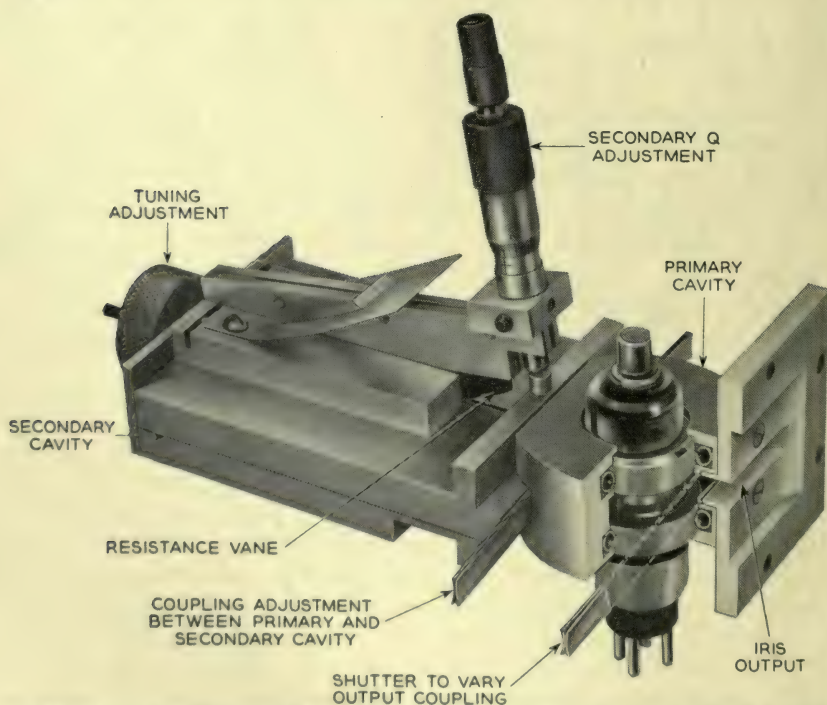


Fig. 14 — An experimental coupled resonator reflex klystron having a fixed frequency primary cavity.

Operation in the 4000-mc band was found to require a closely fitting primary cavity resonating in its principal mode. This unit, made of gold plated brass, is coupled to the output waveguide and secondary cavity respectively through two rectangular irises the sizes of which are independently variable by means of shutters. Toroidal contact springs located in circular grooves grip the inserted tube and complete the external circuit. The secondary cavity consists of a length of rectangular waveguide and uses a movable contacting plunger for tuning. A micrometer driven resistance vane may be inserted into this cavity through a longitudinal slot in its top surface thereby obtaining a wide range of continuously variable Q 's. The entire unit is attached to the output waveguide by means of the adapter plate shown to the right of the primary cavity in Fig. 14.

A coupled resonator circuit having a tunable primary cavity is shown in the photograph of Fig. 15. The secondary cavity of this unit is identical to the one shown in the Fig. 14 but the primary resonator has been

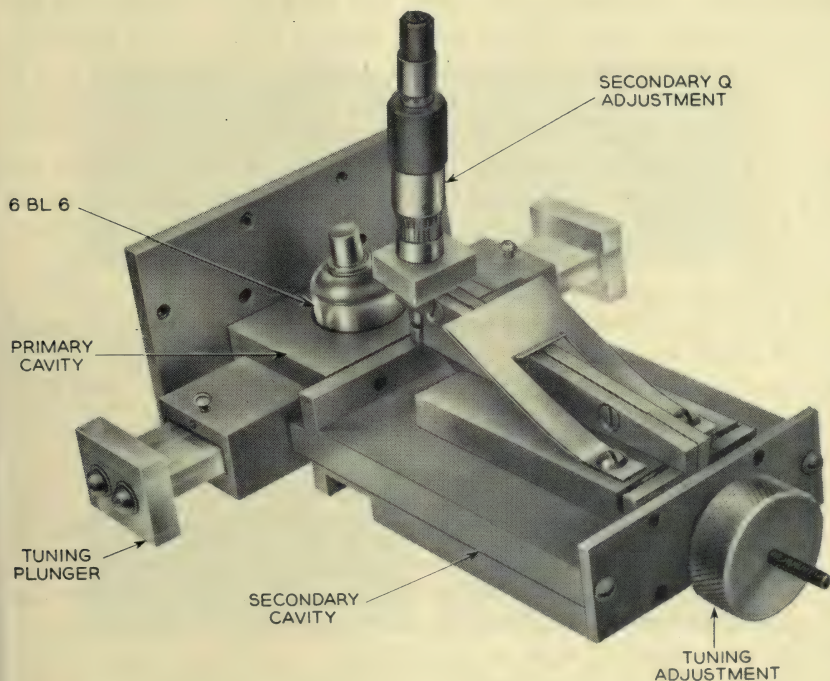


Fig. 15 — An experimental coupled resonator reflex klystron with a tunable primary cavity. The mechanical tuning range of this tube extends from 3700–4200 mc.

modified to take two non-contacting tuning plungers so dimensioned that they can be moved in very close to the glass wall of the 6BL6. In this way an operating frequency range from 3700–4200 mc was obtained.

4.2 Qualitative Verification of Theory

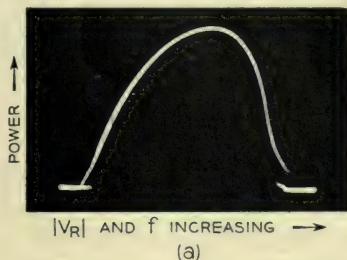
The results of a qualitative verification of the coupled-resonator reflex klystron theory are given in the form of oscillographic displays in Fig. 16. Photograph (a) shows the $(3 + \frac{3}{4})$ -repeller mode of the 6BL6 with the secondary cavity completely decoupled. In photograph (b) the coupling iris and secondary Q have been adjusted for a flat topped output curve. The bandwidth of this curve for a variation in power output of ± 0.1 db was found to be 58 mc, i.e., considerably greater than the half-power bandwidth of the single resonator case above. Increasing Q_k beyond this point by either enlarging the coupling iris or withdrawing the resistance vane from the secondary resonator or a combination of both brings about the condition of critical coupling, illustrated by photograph (c). As explained earlier, this mode-shape results from the admittance plot of the passive circuit forming a cusp in the g - b plane. Comparison between this photograph and the theoretical mode of Fig. 9(b) for $(Qk)^2 = 1$ indicates good qualitative agreement. Coupling the secondary cavity still tighter, i.e., making the value of Q_k greater than unity, causes the formation of a loop in the circuit admittance plot and the consequent load hysteresis shown in photograph (d). Whereas all oscillograms discussed to this point have been obtained with a 60-cycle sinusoidal sweep applied to the repeller, the last one in this group, (e), results from a unidirectional (sawtooth) repeller sweep. The nature of this hysteresis effect has been explained earlier, and oscillograms one might expect to observe with overcoupled resonators for both sinusoidal and unidirectional sweeps were shown in Fig. 10. Again we note good qualitative agreement.

4.3 Quantitative Verification of Theory

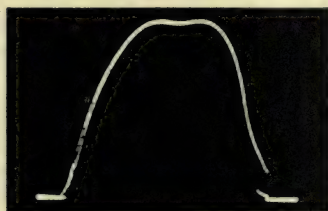
As further strengthening of the theory underlying the operation of the coupled resonator reflex klystron a quantitative, experimental verification was undertaken. The methods used in connection with this work and the results obtained will form the subject of the sections to follow. To anticipate some of the conclusions which will be presented in the course of this description, let it be said here that the quantitative agreement between theory and experiment was found to be of an order high enough to justify amply the approximations involved in the ex-

pressions for both electronic admittance and passive circuit admittance, and to establish confidence in the method of analysis proposed to predict tube performance under a variety of conditions.

Since theoretical results had been worked out for the cases of the secondary Q equal to the primary Q and for the secondary Q equal to



SECONDARY RESONATOR COMPLETELY DECOUPLED. HALF POWER ELECTRONIC TUNING RANGE EQUAL TO 49 MC.
(PEAK POWER = 100 MW)



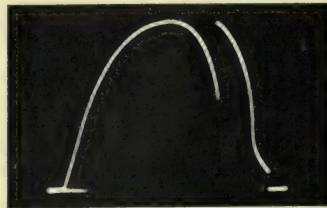
SECONDARY RESONATOR COUPLING AND Q ADJUSTED FOR FLAT TOPPED MODE SHAPE. BANDWIDTH FOR POWER VARIATION OF ± 0.1 DB IS 59 MC.



MODE SHAPE FOR CONDITION OF CRITICAL COUPLING, I.E., $Q_S K = 1$



LOAD Hysteresis DUE TO OVER-COUPLING, I.E., $Q_S K > 1$



SAME AS (d) EXCEPT FOR UNIDIRECTIONAL (SAWTOOTH) REPELLER SWEEP

Fig. 16 — Qualitative verification of theory. These oscillograms were obtained with a 6BL6 in the circuit of Fig. 14 and the following operating conditions: $V_0 = 325V$, $I_k = 28$ ma, $f_0 = 3800$ mc, $3 + \frac{3}{4}$ cycle repeller mode.

half the primary one, (see Figs. 7 and 11, respectively) it was decided to establish the same conditions in the experimental tube and to compare the results thus obtained with those predicted by theory.

Prerequisites for the execution of these experiments were:

- (a) Knowledge of the primary Q .
- (b) Complete calibration of the secondary cavity with respect to both the variation of Q_s with penetration of resistance vane and variation of coupling coefficient with size of coupling iris.
- (c) Knowledge of the location of the load lines with respect to the small signal electronic admittance plot in the complex admittance plane.

4.3.1 *Determination of Primary Q .* The above parameter is understood to denote the "operating" or loaded Q of the tube with the output iris adjusted to its final and permanent size and the secondary cavity completely decoupled. For the case of an inductively tuned (fixed gap) reflex klystron, it may be determined experimentally using the expression,⁷

$$Q = \pi f_0^2 \left| \frac{d\tau}{dV_R} \right| \bigg|_{f\tau=N+3/4} - \pi(N + \frac{3}{4}), \quad (4.1)$$

where τ denotes the repeller space transit time. Both $d\tau/dV_R$ and df/dV_R are to be evaluated at the center of the mode.

The functional relationship between τ and V_R was obtained by placing the 6BL6 in the tunable primary cavity of Fig. 15 and determining the repeller voltage for maximum power output over the mechanical tuning range of the cavity. Since the same repeller mode was used throughout this test, τ must also have been the same at each of these frequencies, namely equal to $(N + 3/4)/f$ seconds. The experimentally determined plot of repeller voltage vs. frequency is given in Fig. 17. It is seen to be a straight line described by the equation,

$$f = (13.5V_R + 2495) \text{ mc} \quad (4.2)$$

and since, for the $3 + 3/4$ -repeller mode,

$$\tau = \frac{3.75}{f_{(\text{mc})}} 10^{-6} \text{ sec.}, \quad (4.3)$$

the desired relation between τ and V_R is obtained from the above two equations as

$$\tau = (3.6V_R + 665)^{-1} \text{ sec.}, \quad (4.4)$$

whence the numerator of equation 4.1 follows as

$$\left| \frac{d\tau}{dV_R} \right| = \frac{3.6}{(3.6V_R + 665)^2} 10^{-6} \text{ sec/volt.} \quad (4.5)$$

It remained only to replace the 6BL6 in the fixed tuned primary cavity, in which all subsequent tests were performed, and to determine the midmode repeller voltage and frequency. These values were found to be 96 volts and 3800 mc respectively, thus yielding by substitution into equation (4.5),

$$\left| \frac{d\tau}{dV_R} \right| = \frac{3.6 \times 10^{-6}}{(3.6 \times 96 + 665)^2} = 3.52 \times 10^{-12} \text{ sec/volt.}$$

The demoninator of equation (4.1) simply denotes the "modulation sensitivity" at the center of the mode. A simple measurement estab-

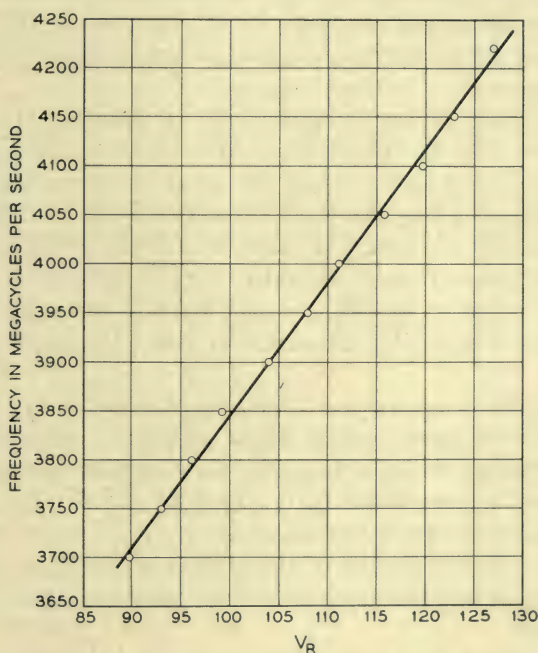


Fig. 17 — Experimental determination of the relation between frequency and repeller voltage. Points shown were obtained with a 6BL6 in tunable primary cavity of Fig. 14 with secondary cavity completely decoupled and operation in the $3 + \frac{3}{4}$ repeller mode at a beam voltage, $V_0 = 325V$. Frequency was varied by means of tuning plungers and repeller voltage recorded for maximum power output. Relation is given by, $f = (13.5V_R + 2495)$ mc.

lished its value as,

$$\left| \frac{df}{dV_R} \right| = 1.12 \text{ mc/volt.}$$

Substitution into equation (4.2) then gave the value of primary Q as,

$$Q = \pi(3.8)^2 10^{18} \frac{(3.52)10^{-12}}{(1.12)10^6} - \pi(3.75) = 130.$$

The entire Q -determination, incidentally, as outlined above, was repeated for the $2 + \frac{3}{4}$ mode and resulted in $Q = 128.8$ thus affording an excellent and independent double check.

4.3.2 Calibration of Secondary Resonator. To facilitate the establishment of controlled and reproducible conditions of secondary Q and coupling coefficient, two calibration curves had to be obtained. One, relating the values of secondary Q with the readings of the micrometer controlling the depth of insertion of the resistance vane and the other, relating the coefficient of coupling, k , with the coupling iris width. The latter could be varied by means of gold plated spring shutters as shown in Figs. 14 and 15.

This calibration of the secondary cavity was carried out in three distinct phases. The first phase involved the determination of the variation of Q_s with micrometer setting for values of Q_s ranging from 500 to 2000. The second phase, which was based on the results obtained in phase one, yielded the complete calibration of the coupling iris and the third and last phase, in turn dependent on results of phase two, yielded values of Q_s down to 65. The reasons for this particular sequence of measurements will become apparent in the following more detailed description.

By means of a Q -measurement technique⁸ based on an oscillographic display of reflected power, points on the calibration curve were obtained as indicated by the circles in Fig. 18. It is seen that the lowest value of Q_s which could be determined by this method was 550. For lower values of Q , the sweep range of the signal generator became insufficient to display the required fraction of the resonance curve; in addition, the cavity proved excessively undercoupled to permit reliable measurements of bandwidths.

Using the values of Q_s thus determined and a particular property of coupled resonators covered earlier in this text and further elaborated below, the relation between coupling coefficient and iris width was established. It was shown earlier that an overcoupled secondary cavity gives rise to load hysteresis manifesting itself in mode shapes as sketched

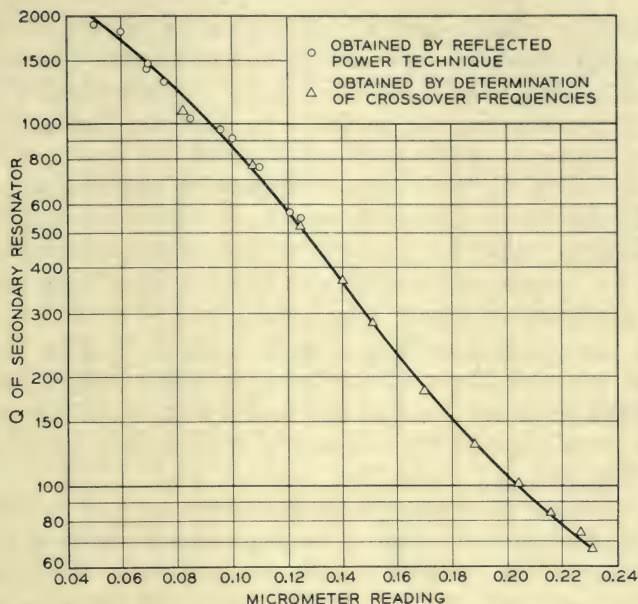


Fig. 18 — Calibration of secondary cavity. Variation of secondary Q with reading of micrometer controlling depth of insertion of resistance vane.

in Fig. 10 and shown in the form of an oscilloscope pattern in Fig. 16(d). Conditions pertaining at the crossover point were described by equation (3.13), which stated:

$$p^2(Qk)^2 = 1 + p^2(2Q\delta)^2,$$

or, since $pQ = Q_s$

$$(Q_s k)^2 = 1 + (2Q_s \delta)^2.$$

Hence the two frequencies corresponding to the crossover point are given by

$$\delta_{1,2} = \pm \frac{1}{2} \sqrt{k^2 - \left(\frac{1}{Q_s}\right)^2} \quad (4.6)$$

and their difference, $\Delta\delta$, by

$$\Delta\delta = \sqrt{k^2 - \left(\frac{1}{Q_s}\right)^2}. \quad (4.7)$$

Equation (4.7)^{4, 5} provides an experimental method of determining k since $\Delta\delta$ can be accurately measured and Q_s is known from previous tests.

The actual procedure adopted in calibrating the coupling iris was this: The secondary cavity was set to a fairly high and known value of Q_s , say 1500, and the coupling iris adjusted to the smallest width which still resulted in overcoupling, i.e. in $Q_s k > 1$, and thus displayed the desired load hysteresis on the oscilloscope screen. By means of a high- Q absorption wavemeter the crossover frequencies were determined and, along with the known value of Q , substituted into equation (4.7) which then yielded the value of k . As a double check the above procedure was repeated for several values of Q_s and in accordance with theory the coupling coefficient, k , found to be a function only of the iris coupling width. A typical set of readings is given in Table II. They show that though Q_s was varied by a ratio of nearly 3:1, the resulting values of k only differed by about one-half per cent. Measurements as shown in the table were repeated for various iris widths. They resulted in the curve of Fig. 19.

To complete the calibration of the secondary cavity it was necessary to extend the relation between Q_s and micrometer readings to values lower than could be obtained by the reflected power technique. This was carried out with the aid of Fig. 19 in the following manner: With the coupling iris adjusted to its maximum width of 0.5", the secondary Q was varied until the condition of critical coupling, as observed by the formation of a cusp on the oscilloscope pattern (see Fig. 16(c)), was reached. Since this condition is characterized by $Q_s k = 1$ and since the value of k corresponding to the iris opening could be read off the curve of Fig. 19, the value of Q_s then followed simply as the reciprocal of k . This measurement was repeated for successively smaller iris openings and the points shown as triangles in Fig. 18 were obtained. Inspection of this figure also reveals that these values of Q_s , ranging from 65 to 1100, overlap and coincide with values of Q_s determined

TABLE II

Coupling Iris Width (inches)	Secondary Cavity Q_s	Crossover Frequencies f_1 and f_2 (mc)	$\Delta f =$ $(f_1 - f_2)$ (mc)	$\frac{\Delta \delta =}{2\Delta f}$ $f_1 + f_2$	From Equation (4.7), $k = \frac{1}{\sqrt{(\Delta \delta)^2 + (1/Q_s)^2}}$
0.500	1430	3749.9 3694.4	55.5	14.92×10^{-3}	14.93×10^{-3}
0.500	975	3748.4 3693.1	55.3	14.90×10^{-3}	14.90×10^{-3}
0.500	573	3745.2 3690.4	54.8	14.74×10^{-3}	14.85×10^{-3}

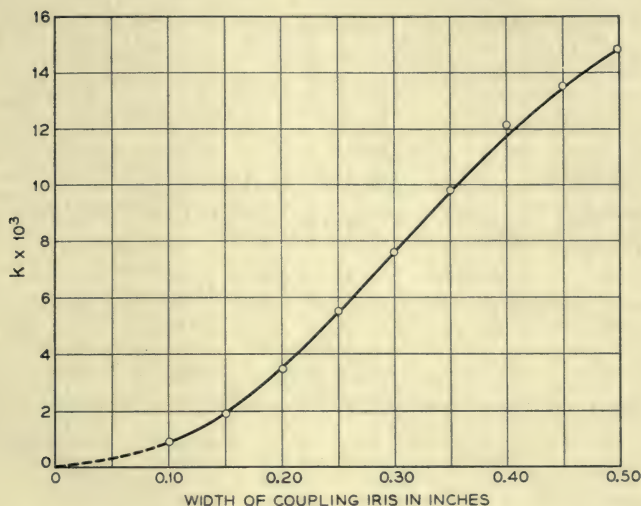


Fig. 19 — Calibration of coupling iris. Variation of coupling coefficient with width of coupling iris.

independently by the reflected power method. The significance of this experimental agreement is twofold. It serves to prove the correctness of the calibration curves of Figs. 18 and 19 and beyond that may be regarded as verifying much of the theory of the coupled resonator reflex klystron.

4.3.3 Comparison of Experimental and Theoretical Mode Shapes. Knowing the primary (operating) Q and the calibration of the secondary resonator, controlled operating conditions were established and the experimental mode shapes shown in the left hand column of Fig. 20 obtained. The first three families of curves in this column are the results of point by point measurements in which repeller voltage, taken as the independent variable, was varied by known amounts about its midmode value and the corresponding values of RF power output and frequency were determined by a thermistor-bridge-wattmeter and high- Q wave-meter respectively. The last family of curves in this group, (d), is derived from (c); in it power has been normalized with respect to its maximum value within the mode thus permitting the convenient determination of midmode percentage reduction of power and electronic tuning range to any desired power level for various values of $(Qk)^2$.

Additional details of the test conditions which gave rise to the ex-

$V_0 = 325$ V, $I_K = 28$ MA, V_R (MIDMODE) $= -96$ V,
 $f_0 = 3800$ MC, $3\frac{3}{4}$ REPELLER MODE,
 PRIMARY $Q = 130$, SECONDARY $Q = 65$,
 Qk VARIED BY CHANGES IN COUPLING IRIS WIDTH

OBTAINED GRAPHICALLY FROM
 COMPLEX ADMITTANCE PLANE
 REPRESENTATION. NORMALIZED FREQUENCY
 $Q\delta$, CONVERTED TO MC BY ASSUMING
 $Q = 130$ AND $f_0 = 3800$ MC

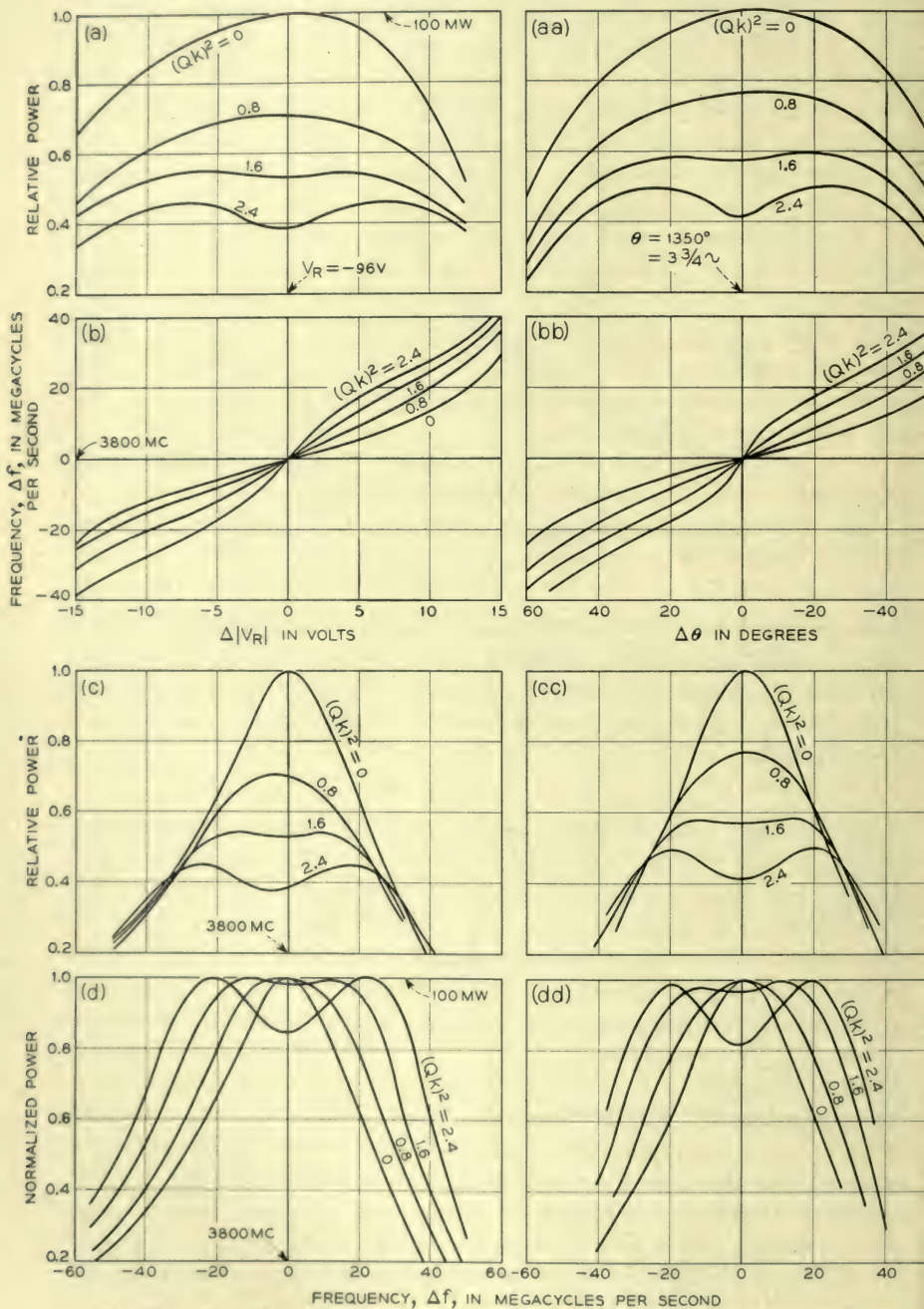


Fig. 20 — Quantitative verification of coupled resonator reflex klystron theory. Curves (a) and (b) were obtained experimentally by changing the repeller voltage about its midmode value and noting the corresponding values of power and frequency. Curves (c) and (d) were deduced from (a) and (b). The theoretical curves were obtained graphically from a complex admittance plane plot similar to those of Figs. 9 and 13.

perimental curves shown are as follows:

6BL6 repeller mode = $3 + \frac{3}{4}$ cycles

Resonator voltage, $V_0 = 325$ volts

Cathode current, $I_k = 28$ ma

Midmode repeller voltage, $V_R = -96$ volts

Midmode frequency, $f_0 = 3800$ mc

Secondary Q , $Q_s = \frac{1}{2}$ primary $Q = 65$, obtained by micrometer reading of 0.234 (see Fig. 18)

The parameter, $(Qk)^2$, was varied by adjusting the iris width, and hence k , according to Table III.

A few words, next, about the method by which the theoretical curves of Fig. 20 were derived. Input admittance plots for two coupled resonators having $Q_s = \frac{1}{2}Q$ were given earlier in this report (see Fig. 11).

TABLE III

$(Qk)^2$	k (for $Q = 130$)	Iris width from Fig. 19 in inches
0.0	0.00	0.000
0.8	6.88×10^{-3}	0.285
1.6	9.73×10^{-3}	0.350
2.4	11.90×10^{-3}	0.405

These included graphs for the values of $(Qk)^2$ listed in Table III and thus of direct use in a graphical admittance plane analysis such as described in Section 3.2. To establish correlation, however, between theory and experiment, it was necessary to determine the location of the load lines in relation to the small-signal electronic admittance plot for the $3 + \frac{3}{4}$ repeller mode. This was performed graphically by trial and error as follows: a number of electronic admittance loci were drawn and their interactions with the load line for $(Qk)^2 = 2.4$ studied. In particular, the ratio of minimum to maximum power was calculated for each trial until one was found agreeing well with the experimental value of 0.845 (see Fig. 20d). This latter trial, then, was taken as representing the proper relation between electronic and circuit admittance. Mode plots for other values of $(Qk)^2$ were determined without further reference to the experimental results and allowed to fall where they may.

In comparing the theoretical and experimental curves, then, the following points should be kept in mind: the frequency scales of the theoretical curves are normally expressed in terms of $Q\delta$. This was converted

to megacycles in Fig. 20 by using the relation,

$$\Delta f = (Q\delta) \frac{f_0}{Q} \text{ mc,}$$

where Q was found to equal 130 by the independent measurement described in section 4.3.1 and f_0 taken as 3800 mc. The vertical power scales of the theoretical curves are not entirely independent of the experimental plots in as much as one of the latter, namely the plot for $(Qk)^2 = 2.4$, was used in determining the location of the load lines within the spiral diagram.

The curves presented in Fig. 20 are largely self-explaining. They indicate good agreement between theory and experiment. What disagreement there is, may be traced primarily to the assumptions involved in the small signal klystron theory not being fully met in practice. Thus, the slight discrepancy between Figs. 20(a) and (aa) may be due to the drift angle not being linearly related to repeller voltage or possibly due to the phase angle of the electronic admittance being affected by the magnitude of the gap voltage.¹¹ These factors, however, are eliminated in the mode plots (c) and (d), which for this reason exhibit better mode symmetry and very close agreement with theory.

4.4 Performance Data

The experimental curves of Fig. 20 were obtained with particular values of $(Qk)^2$ and Q_s/Q for which the input admittance of two coupled resonators had been computed and plotted earlier. As pointed out before, these values were chosen merely because they facilitated comparison between theory and experiment; they were not to be regarded, however, as representing optimum conditions. Whereas these curves had been obtained for a fixed secondary Q (namely equal to half the primary one), with changes in coupling-iris width producing the desired variations in $(Qk)^2$, the demonstration of optimum performance which follows was pursued along different lines. Here, the coupling iris was opened to its maximum width (0.500") and the secondary Q adjusted until the conditions shown by the oscillograms of Fig. 21 were obtained.* These oscillograms indicate a number of mode shapes useful in applications requiring power output to be essentially independent of frequency. It is

* In adjusting the circuit parameters for a flat topped mode it should be borne in mind that the variation in power with frequency is a function of $Q_s k$ whereas the actual bandwidth varies inversely with Q_s . For a maximum flat band, therefore, Q_s should be chosen as small as possible consistent with a value of $(Q_s k)^2$ of about 0.35. In practice the lowest value of Q_s which can be used will be determined by the highest value of k obtainable with a given coupling iris.

seen that the useable bandwidth depends on the degree of flatness. Thus, if the application requires power to be absolutely constant, the useable bandwidth equals 39 mc; it increases to about 70 mc for a fluctuation in power of ± 0.2 db. Fig. 21 also lists the values of half power electronic tuning for the oscillograms shown. They are seen to range between 107 and 113 mc.

By way of comparison, the mode shape for the case of a completely decoupled secondary resonator was shown in Fig. 16(a). Its peak power was found to equal about 100 mw, or about twice the power of the flat topped modes of Fig. 21, and its half power electronic range equaled 50

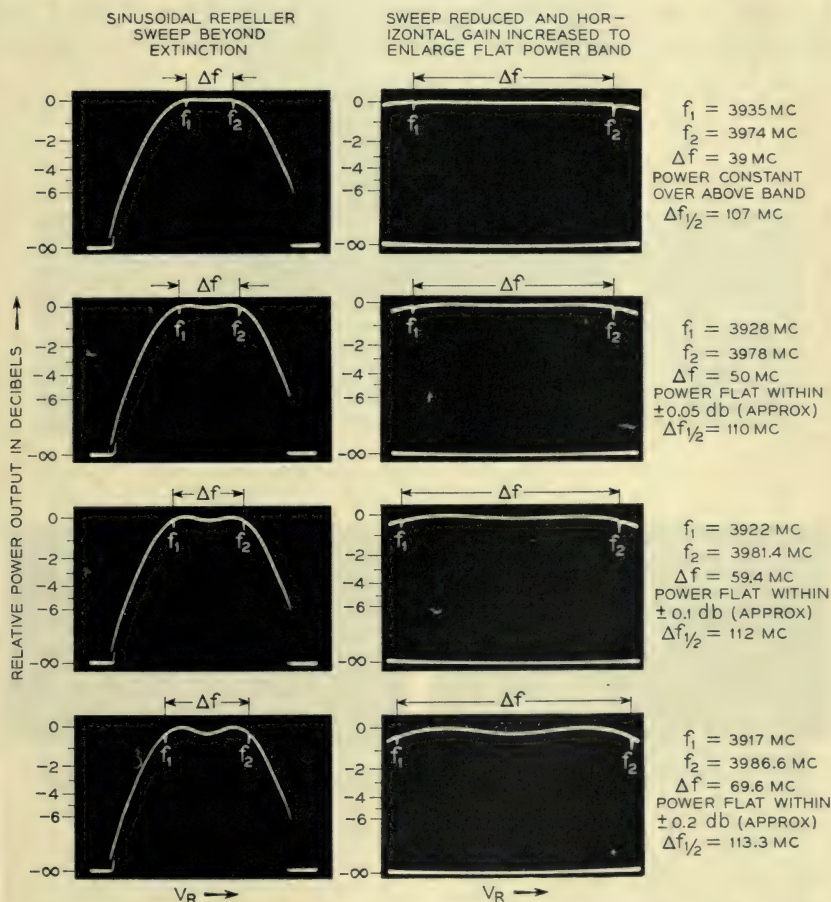


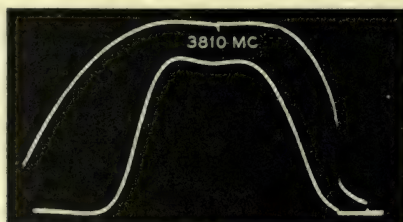
Fig. 21 — Oscillograms showing flat power bands obtainable with coupled resonator reflex klystron. 6BL6 operating in $3 + \frac{3}{4}$ mode, $V_0 = 325$ V, $I_k = 28$ ma, zero DB power level corresponds to 50 milliwatts.

mc. The price, then, we must pay for a better than two fold increase in electronic tuning and for a frequency range of constant power is a 3 db loss in the available power output.

Using the tunable primary cavity shown in Fig. 15, the mode performance just described could be obtained about any center frequency between 3700 and 4200 mc.

5.0 APPLICATIONS OF THE COUPLED-RESONATOR REFLEX KLYSTRON

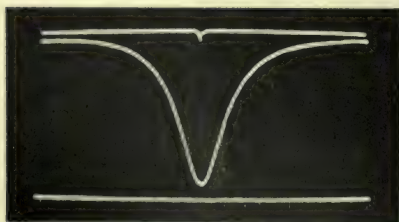
A number of applications making use of the variety of mode shapes which may be obtained with the coupled resonator reflex klystron were indicated in the introductory section of this paper. Some of these applications were tried experimentally and are described below, others,



(a)

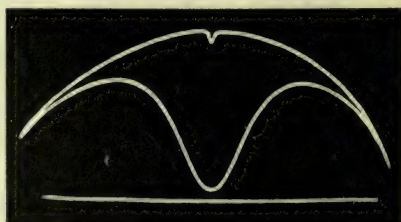
USE OF COUPLED-RESONATOR REFLEX KLYSTRON IN GAIN COMPARATOR TO DISPLAY BANDPASS CHARACTERISTIC OF MICROWAVE CHANNEL FILTER. TOP TRACE REPRESENTS INCIDENT POWER, LOWER TRACE TRANSMITTED POWER.

CENTER FREQUENCY OF FILTER = 3810 MC
PASS BAND = 28 MC



(b)

USE OF COUPLED-RESONATOR REFLEX KLYSTRON AS SWEEPED FREQUENCY SOURCE IN DETERMINATION OF Q OF UNKNOWN RESONATOR BY REFLECTED POWER METHOD. TOP TRACE REPRESENTS POWER INCIDENT UPON CAVITY UNDER TEST, MIDDLE TRACE REFLECTED POWER AND LOWER TRACE ZERO POWER LEVEL. (NOTE FLATNESS OF INCIDENT POWER TRACE)



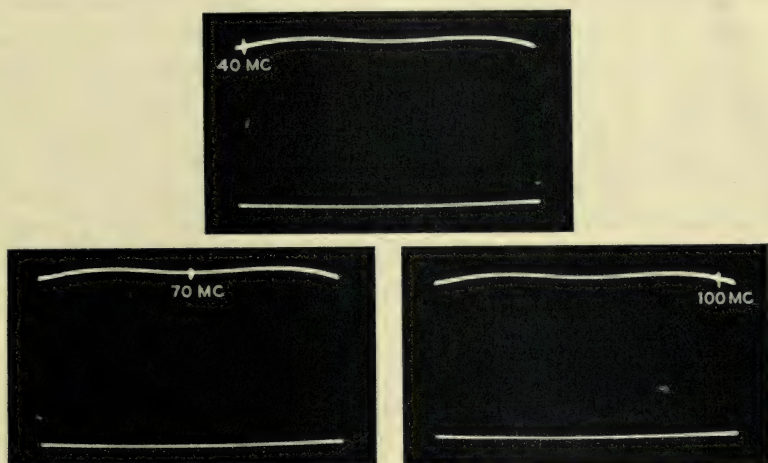
(c)

OSCILLOSCOPE DISPLAY FOR SAME RESONATOR RESULTING FROM USE OF CONVENTIONAL KLYSTRON AS SWEEPED FREQUENCY SOURCE

Fig. 22 — Examples of application of coupled resonator reflex klystron as microwave sweeper.

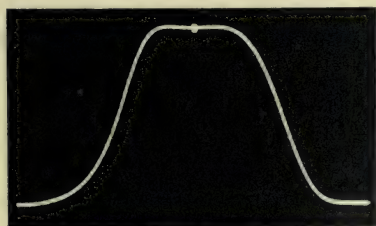
and in particular those making use of the improved modulation linearity resulting from the use of two resonators, could not be tried since the required test apparatus⁹ was unavailable. The close correlation between theory and experiment, however, as demonstrated in Section 4.3.3 leaves little doubt as to the feasibility of such applications.

Two examples of the use of the coupled resonator reflex klystron as a microwave sweeper are given in Fig. 22. The first example shows the band pass characteristic of a channel filter, 28 mc wide. The second example demonstrates the use of this tube in a Q -measurement scheme⁸



(a)

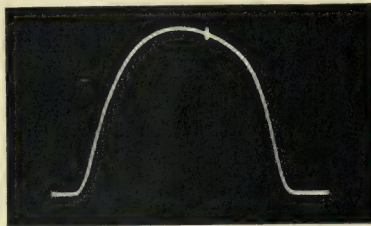
SUCCESSIVE PHOTOGRAPHS OF OSCILLOSCOPE DISPLAY
SHOWING IF-POWER AS A FUNCTION OF FREQUENCY WITH
SIGNAL GENERATOR TUNED TO DIFFERENT FREQUENCIES



(b)

GAIN-FREQUENCY CHARACTERISTIC OF
IF AMPLIFIER USING COUPLED-RESONATOR
REFLEX KLYSTRON AS SWEEP SOURCE

CENTER FREQUENCY OF IF-AMPL = 70MC
FLAT BANDWIDTH OF IF-AMPL = 22 MC



(c)

GAIN-FREQUENCY CHARACTERISTIC OF
SAME IF AMPLIFIER AS OBTAINED BY
CONVENTIONAL (SINGLE RESONATOR)
REFLEX KLYSTRON. DISTORTION IN GAIN
ENVELOPE DUE TO MODE SHAPE OF
KLYSTRON.

Fig. 23 — Example of application of coupled resonator reflex klystron as intermediate frequency sweeper.

based on the oscillographic display of reflected power. The advantage of a flat-topped mode shape for this measurement may be seen by comparison with the oscillogram of Fig. 22(c) which was obtained for the same resonator under test using a conventional reflex klystron as swept frequency source.

The center frequency of the flat power band obtainable with the coupled resonator reflex klystron may be shifted into the intermediate frequency range by mixing its output with a suitable local oscillator. This is shown in Fig. 23 where the first three oscillograms represent a power band flat to within ± 0.1 db and 60 mc wide centered around an IF frequency of 70 mc. The frequency markers are obtained from a signal generator tuned to the frequencies indicated and coupled lightly to the output of the mixer. Fig. 23(b) shows the gain-frequency characteristic of a 70 mc-IF strip having a useful bandwidth of 22 mc. The last oscillogram shows the gain-frequency characteristic of the same IF amplifier with the coupled resonator reflex klystron replaced by one of conventional design. Due to the inadequate electronic tuning range of this klystron and due to the asymmetry of its mode shape, the gain envelope of the IF amplifier appears distorted.

6.0 CONCLUSIONS

The theory and reduction to practice of a coupled resonator reflex klystron have been presented. This device differs from the conventional reflex klystron in that the electronic admittance interacts with the input admittance of two coupled resonators. Significant and advantageous changes in performance resulting therefrom are:

(1) Power output can be made to be substantially flat over part of the electronic tuning range. The frequency range of flat power is greater than the half power electronic tuning range of a klystron having the same electron-optical system but interacting with a single resonator.

(2) The half-power electronic tuning range of the coupled resonator klystron is more than twice that of the equivalent single resonator klystron.

(3) Modulation linearity may be obtained over a greatly increased frequency swing.

A reduction in power output of about 3 db occurs for a secondary resonator Q and coupling coefficient adjustment designed to yield a maximum flat band or maximum electronic tuning, a much smaller reduction in output power, perhaps 10 per cent, will provide a substantial improvement in modulation linearity.

Qualitative and quantitative experimental verifications of the theory were undertaken and good agreement obtained. Oscillograms of mode shapes having a range of constant power output and wide half-power electronic tuning were presented together with a number of applications of the coupled resonator reflex klystron both as a microwave and as an intermediate frequency sweep source. Other applications were indicated and additional ones may suggest themselves to the reader.

It is believed that the addition of the second resonator makes the reflex klystron a more useful device in many of its more important applications.

ACKNOWLEDGEMENTS

The writer wishes to express his appreciation to J. O. McNally for the advice and encouragement received in the execution of this work. The assistance of the thesis advisor, Prof. John B. Russell of the Electrical Engineering Department of Columbia University, is also gratefully acknowledged.

REFERENCES

1. Pierce, J. R., and W. G. Shepherd, Reflex Oscillators. *Bell System Tech. J.*, **26**, pp. 460-690, July, 1947. This reference constitutes a very complete treatment of reflex klystron theory and practice. The derivation of the expression for electronic admittance is given in Appendix III, pages 639-643; of particular interest to the coupled resonator reflex klystron is Pierce and Shepherd's treatment of Frequency Sensitive Loads and Long Lines Effects, pp. 523, where it is shown that certain load conditions encountered in the operation of reflex klystrons may result in load hysteresis similar to the one described in the preceding paper as due to an overcoupled secondary resonator.
2. Martin, R. A., and R. D. Teasdale, Input Admittance Characteristics of a Tuned Coupled Circuit. *I. R. E., Proc.*, p. 57, January, 1952, and correction to this paper published in *I. R. E., Proc.*, p. 459, April, 1952. This paper presents a precise expression for the input admittance of two coupled resonant circuits and gives plots of specific numerical examples. The expression for input admittance given in the above named correction reduces to equation (3.2) and (3.12).
3. Bleaney, B., Electronic Tuning of Reflection Klystrons. *Wireless Engineer*, p. 6, Jan., 1948. A valuable background paper on the mechanics and maximization of electronic tuning.
4. Very High-Frequency Techniques, Vol. II, Radio Research Laboratory, Harvard University. (McGraw-Hill). Chapter 31, p. 849 contains a concise and valuable treatment of reflex klystron theory. It leads up to the application of external-cavity reflex klystrons in wide tuning range coaxial resonators. In the course of explaining some load hysteresis phenomena which are sometimes caused in these circuits by parasitic resonances or excitation of undesired modes, a simplified expression for the input admittance of two coupled resonators is given. This expression is similar to equation (3.12). An experimental method is also suggested (p. 869) to measure the coupling coefficient between the main resonator and the undesired resonance. Use has been made of this method in the calibration of the secondary resonator in Section 4.3.2.

5. Spangenberg, K. R., *Vacuum Tubes*, p. 601. (McGraw-Hill). This reference covers essentially the same material as Reference 4.
6. Reference Data for Radio Engineers, Third Edition, p. 121, Federal Telephone and Telegraph Corp. Here a universal chart of the selectivity (transfer characteristic) of two coupled tuned circuits is presented. The format and choice of variables of this presentation has inspired the selection of the equivalent quantities applying to the driving point properties of coupled tuned circuits.
7. Harmen, W. W., and J. H. Tillotson, Beam-Loading Effects in Small Klystrons. I. R. E., Proc., p. 1419, Dec., 1949. Equation (3) of this reference is identical with equation (4.1) and has been used in the experimental determination of the primary operating Q .
8. Reed, E. D., A Sweep Frequency Method of Q Measurement for Single-Ended Resonators. Proc. National Electronics Conference, Vol. VII, p. 162. Also reprinted as Bell Telephone System Monograph 1953. This method of Q -measurement was used in the calibration of the secondary resonator described in Section 4.3.2. The coupled resonator reflex klystron provides a useful tool for this type of measurement as may be seen from Fig. 22(b).
9. Albersheim, W. J., Measurement Techniques for Broad-Band Long-Distance Radio Relay Systems. I.R.E., Proc., p. 548, May, 1952. Figs. 4(a) and 4(b) of this reference give the block diagram of a "Linearity-Test-Set" developed at Bell Telephone Laboratories to determine the modulation linearity of a reflex klystron used as FM-deviator.
10. Boehlke, P. G., and F. C. Breedan, External Cavity Klystron. Electronics, p. 114, July, 1947. This paper gives an account of the design considerations which led to the development of the 6BL6 reflex klystron. It also contains useful information on its performance characteristics.
11. Garrison, J. B., A Qualitative Analysis of Hysteresis in Reflex Oscillators. Radiation Laboratory Report No. 650, dated Feb. 4, 1946. In this treatment on the causes of electronic hysteresis Garrison shows (on page 5) that a dependence of the phase angle of the electronic admittance upon the magnitude of RF gap voltage will give rise to asymmetry in the mode plot of power output vs. repeller voltage although the power vs. frequency relation will be symmetrical. (Note that the experimental mode plots for $Qk = 0$ of Figs. 20(a) and (c) show evidence of the same phenomenon.)

Abstracts of Bell System Technical Papers* Not Published in This Journal

ANDERSON, P. W.¹

Concept of Spin-Lattice Relaxation in Ferromagnetic Materials, Letter to the Editor, *Phys. Rev.*, **88**, pp. 1214, Dec. 1, 1952.

BALASHEK, S., see K. H. Davis.

BAND, W., see R. A. Nelson.

BEACH, A. L., see J. J. Lander.

BELL TELEPHONE LABORATORIES Transistor Teachers Summer School.

Experimental Verification of the Relationship Between Diffusion Constant and Mobility of Electrons and Holes, *Phys. Rev.*, **88**, pp. 1368-1369, Dec. 15, 1952.

The relationship between diffusion constant and mobility, called the Einstein relationship has been experimentally verified for electrons and holes in germanium. This has been accomplished by measuring the rate of increase in half concentration width of a pulse of minority carriers moving in an electric field.

BIDDULPH, R., see K. H. Davis.

BRIDGMAN, D. C.²

College Graduates and the Country's Telephone Industry, *Jl. College Placement*, **13**, pp. 19-27, Oct., 1952.

BULLINGTON, K.¹

Radio Transmission Beyond the Horizon in the 40- to 4,000-Mc Band, *I.R.E., Proc.*, **41**, pp. 132-135, Jan., 1953. (Monograph 2060).

Reliable signals have been received at distances of several hundred miles at frequencies of 500 and 3,700 mc. The median signal levels are 50 to 90

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

¹ Bell Telephone Laboratories, Inc.

² American Telephone and Telegraph Company.

db below the free-space field, but are hundreds of db (in one case 700 db) stronger than the value predicted by the classical theory based on a smooth spherical earth with a standard atmosphere. Antenna gains and beam widths are maintained to a first approximation and no long delayed echoes have been found.

CLARK, A. B.¹

Development of Telephony in the United States, A.I.E.E., Trans., Commun. and Electronics Sect., **71**, pp. 348-364, Nov., 1952 (Monograph 2045).

The telephone was invented twenty-four years after the founding of the American Society of Civil Engineers and Architects, which society commemorated its Centennial in September, 1952. The telephone was eight years old when the American Institute of Electrical Engineers was founded. Since Bell's invention, the telephone business has grown tremendously and this growth has been greatly dependent on developments in science and engineering. This paper traces, and endeavors to give the significance of, the major developments. Brief mention is made of some developments still in the making and of some ideas as to the future potentialities of the business.

CONWELL, E. M.¹

Mobility in High Electric Fields, Phys. Rev., **88**, pp. 1379-1380, Dec. 15, 1952.

An extension of conductivity theory to high fields, subject to the usual simplifying assumptions, is carried out for the cases in which the change of energy of an electron in a collision can be neglected. This yields a relationship between mobility and relaxation time which is valid over a wide range of fields.

DAVIS, K. H.,¹ R. BIDDULPH¹ AND S. BALASHEK¹

Automatic Recognition of Spoken Digits, J. Acoust. Soc. Am., **24**, pp. 637-642, Nov., 1952.

The recognizer discussed will automatically recognize telephone-quality digits spoken at normal speech rates by a single individual, with an accuracy varying between 97 and 99 per cent. After some preliminary analysis of the speech of any individual, the circuit can be adjusted to deliver a similar accuracy on the speech of that individual. The circuit is not, however, in its present configuration, capable of performing equally well on the speech of a series of talkers without recourse to such adjustment. Circuitry involves division of the speech spectrum into two frequency bands, one below and the other above 900 cps. Axis-crossing counts are then individually made of both band energies to determine the frequency of the maximum syllabic rate energy within each band. Simultaneous two-dimensional frequency portrayal is found to possess recognition significance. Standards are then determined, one for each digit of the ten-digit series, and are built into the recognizer as a form of elemental memory. By means of a series of calculations performed

¹ Bell Telephone Laboratories, Inc.

on the spoken input digit, a best match type comparison is made with each of the ten standard digit patterns and the digit of best match selected.

DICKTEN, E., see R. L. Wallace, Jr.

FELKER, J. H.¹

Typical Block Diagrams for a Transistor Digital Computer, Elec. Eng., **71**, pp. 1103-1108, Dec., 1952 and A.I.E.E. Trans., **71**, pp. 175-182, 1952 (Monograph 2046).

The superior speed capabilities of vacuum tubes have led to their use in computer designs to replace relays. Because of their small size, low power consumption, and long life expectancy, it now appears that transistors will replace tubes as computer elements. Here is a study of binary computer functions in which transistors are employed.

FRAYNE, J. G.,⁶ AND J. P. LIVADARY⁷

Dual Photomagnetic Intermediate Studio Recording, S.M.P.T.E., J., **59**, pp. 388-397, Nov., 1952.

Selected production magnetic tracks are transferred to a recorder which lays down collinear 200-mil push-pull direct-positive variable-area and magnetic tracks. Magnetic stripe is on base of photosensitive emulsion on the opposite edge of film from photo track. The photo track may be used for reviewing, cutting, etc. Re-recording is done from assembled magnetic tracks. This method combines advantages of photo track for editing and provides superior quality of magnetic track.

HAGSTRUM, H. D.¹

Electron Ejection from Mo by He^+ , He^{++} , and He_2^+ , Phys. Rev., **89**, pp. 244-255, Jan. 1, 1953.

Total yield and kinetic energy distribution have been measured for electrons ejected from atomically clean and gas covered molybdenum by the ions He^+ , He^{++} , and He_2^+ , in the kinetic energy range 10 to 1000 ev. Evidence is presented that one electron is excited into the kinetic energy continuum for each incident. He^+ ion and that the electrons so excited are partially internally reflected at the potential barrier of the metal. The slowest ions observed were found to eject 0.25, 0.72, and 0.13 electron per ion for He^+ , He^{++} , and He_2^+ , respectively. Total electron yield is found to be nearly independent of ion kinetic energy up to 1000 ev. This observation and that of the kinetic energy maximum for slow ions indicate that the electrons are released in an Auger type process for which the energy is supplied by the potential and not the kinetic energy of the ion (potential ejection). Electrons of kinetic energy greater than the upper limit predicted by present theory are observed for faster ions and are accounted for by the shift of the energy levels of the bombarding particle when it is near the metal surface. Some conclusions con-

¹ Bell Telephone Laboratories, Inc.

⁶ Westrex Corporation.

⁷ Columbia Pictures Corporation.

cerning reflection processes at a metal surface and the nature of electron ejection by the alpha-particle (He^{++}) and the molecular helium ion (He_2^+) come out of this work.

HEIDENREICH, R. D.¹

Methods in Electron Microscopy of Solids, Rev. Sci. Instr., **23**, pp. 583-594, Nov., 1952 (Monograph 2047).

Methods of replicating solid surfaces for electron microscopy are reviewed and compared. Preparation of metal surfaces for electron microscopy is discussed, and the advantages of employing electron diffraction techniques in evaluating prepared surfaces are pointed out. Examples of the application of replicas include steel, precipitation in alloys, such as Alnico 5, and studies of slip in aluminum. Growth spirals on the surfaces of crystals of *n*-paraffins are demonstrated. The use of thin metal sections and of emission electron microscopy in studying metallic structures is discussed, and examples are given.

HUTCHINSON, A. R.³

How to Conceal Telephone Wires, Keep Desks Neat, Standardization, **23**, p. 407, Dec., 1952.

JAKES, W. C., JR.¹

Theoretical Study of an Antenna-Reflector Problem, I.R.E., Proc., **41**, pp. 272-274, Feb., 1953.

This paper gives the results of a theoretical investigation of an antenna used with a plane reflector. This finds application in microwave relay stations, where the antenna is placed at ground level facing up and the reflector is located some distance above it. The results given show that there are certain values of λ , separation distance, reflector and antenna size for which the received power is greater than for the same antenna alone at the elevated location.

KAPLAN, E. L.¹

Tensor Notation and the Sampling Cumulants of *k*-Statistics, Biometrika, **39**, pp. 319-323, Dec., 1952.

Now and then in the literature one finds results relating to multivariate distributions which are derived virtually independently of, or with considerable effort from, the corresponding univariate relations, whereas they are in fact only very mild generalizations of the latter, as will be shown. Only the familiar concepts of moments, characteristic functions, cumulants, and *k*-statistics and their sampling cumulants will be discussed here. It should be emphasized that these concepts are identical with those ordinarily used in multivariate situations; the only novelty lies in the concise manner of representing and handling them.

¹ Bell Telephone Laboratories, Inc.

³ Western Electric Company, Inc.

KELLER, A. C.¹

Economics of High-Speed Photography, S.M.P.T.E., Jr., **59**, pp. 365-368, Nov., 1952 (Monograph 2052).

The economics of the use of high-speed photography in research and development work are discussed. High-speed photography is a relatively new tool for engineers which can be used to measure mechanical or electrical effects or both at the same time. Examples are given which illustrate the savings in engineering manpower as well as in materials, devices and systems.

KERN, H. E., see J. J. Lander.

KOCK, W. E.,¹ AND R. L. MILLER.¹

Dynamic Spectrograms of Speech, Letter to the Editor, J. Acoust. Soc. Am., **24**, pp. 783-784, Nov., 1952.

KOCK, W. E.¹

Problem of Selective Voice Control, J. Acoust. Soc. Am., **24**, pp. 625-628, Nov., 1952 (Monograph 2048).

The development of devices which can be operated automatically from the phonetic content of speech may be viewed in terms of the more general problem of the reduction of channel capacity in communications systems. Significance has been observed in formant positions and movements as regard the identification of speech sound. The basic problems in the derivation of phonemes from the formant patterns are reviewed.

LANDER, J. J.,¹ H. E. KERN¹ AND A. L. BEACH¹

Solubility and Diffusion Coefficient of Carbon in Nickel: Reaction Rates of Nickel-Carbon Alloys with Barium Oxide, J. Appl. Phys., **23**, pp. 1305-1309, Dec., 1952.

Experimental values for the solubility of carbon in nickel in the range 700°C to 1300°C yield the equation $\ln S = 2.480 - 4,880/T$, where S is the solubility in grams of carbon per 100 grams of nickel. Values obtained for the diffusion coefficient in the same range fit the equation $\ln D = 0.909 - 20,200/T$, where D is in cm^2 per second. These results are of some interest in the problem of the activation of thermionic oxide coated cathodes, and the experimental method used to measure the diffusion coefficients is related to phenomena occurring in vacuum tubes. To extend the usefulness of the results in this direction, rates of reaction between diffused carbon and barium oxide coatings on nickel have been measured. It was found that the rates are diffusion limited over a wide range of conditions of interest.

LEWIS, H. W.¹

Multiple Meson Production in Nucleon-Nucleon Collisions. Revs. Modern Phys. **24**, pp. 241-248, Oct. 1952 (Monograph 2049).

LIVADARY, J. P., see J. G. Frayne

¹ Bell Telephone Laboratories, Inc.

LUKE, C. L.¹

Photometric Determination of Silicon in Ferrous, Ferromagnetic, Nickel, and Copper Alloys — A Molybdenum Blue Method, Anal. Chem., **25**, pp. 148-151, Jan., 1952.

A simple, rapid photometric method for the determination of silicon in ferrous, ferromagnetic, nickel, and copper alloys has been developed. Wide applicability is its most unique and important feature. Interfering elements are removed by a carbamate-chloroform extraction and silicon is then determined by the photometric molybdenum blue method. Confirmatory data on a wide variety of samples of known silicon content are presented.

LUMSDEN, G. Q.¹

A Quarter Century of Evaluating Pole Preservatives. Amer. Wood Preservers' Assoc., Proc., **48**, pp. 27-47, 1952 (Monograph 1999).

MACWILLIAMS, W. H., JR.¹

Computers — Past, Present, and Future, Elec. Eng., **72**, pp. 116-121, Feb., 1953.

This article deals with the historical development of computers. It also discusses current problems and indicates future structural and functional computer trends which will help to free man from burdensome calculations and increase his material wealth while permitting him more time for pursuits not directly concerned with earning a living.

MILLER, R. L., see W. E. Koch.

MUMFORD, W. W.¹

Optimum Piston Position for Wide-Band Coaxial-to-Waveguide Transducers² I.R.E., Proc., **41**, pp. 256-261, Feb., 1953.

A coaxial line can be matched to a waveguide by means of a probe antenna located ahead of a short-circuiting plunger. An impedance match can usually be achieved by varying any two of the following three dimensions: (a) the off-center position of the probe, (b) the probe length, (c) the piston position. This paper points out that there is, theoretically, an optimum piston position for greatest bandwidth, and presents some evidence corroborating this theory. Bandwidths greater than ± 10 per cent to the 1 db swr points have been realized by fixing the piston at its optimum position and varying (a) and (b) above to obtain a match.

NELSON, R. A.,¹ AND W. BAND⁴

Vapor Pressure of He³ = He⁴ Mixtures, Letter to the Editor, Phys. Rev., **88** p. 1431, Dec. 15, 1952.

¹ Bell Telephone Laboratories, Inc.

² American Telephone and Telegraph Company.

⁴ State College of Washington.

OSBORNE, H. S.²

A Rose by Any Other Name, Report on Work of the Anglo-American Committee on Technical Terminology, Standardization, **24**, pp. 19-20, Jan., 1953.

PECK, D. S.¹

Ten-Stage Cold-Cathode Stepping Tube, Elec. Eng., **71**, pp. 1136-1139, Dec., 1952 (Monograph 2054).

Developments in the art of transferring a gas discharge from one point to another in a multi-electrode tube have led to the design of a 10-stage counting tube operating up to about 2,000 pulses per second. Such a tube can be used for pulse counting, frequency division, time measurements, and similar functions.

PETERSON, G. E.¹

Information-Bearing Elements of Speech, J. Acoust. Soc. Am., **24**, pp. 629-637, Nov., 1952.

This study deals with those aspects of speech which are phonetically significant. A technique has been developed with which phonetically equivalent speech samples may be obtained in different phonetic contexts and from different speakers. Data on two front vowels by different types of speakers are presented. The technique has also been applied to the evaluation of words containing these two vowels.

PIERCE, J. R.¹

New Method of Calculating Microwave Noise in Electron Streams, I.R.E., Proc., **40**, pp. 1675-1680, Dec., 1952.

The noise in a temperature-limited electron beam in a drift space is calculated by a new means. Noise maxima and minima are found. The results agree with calculations made by the Rack-Llewellyn-Peterson method.

RICE, S. O.¹

Statistical Fluctuations of Radio Field Strength Far Beyond the Horizon, I.R.E., Proc., **41**, pp. 274-281, Feb., 1953.

When a sinusoidal radio wave of extremely high frequency is sent out by a transmitter, the wave received far beyond the horizon is often observed to fluctuate. Here some of the statistical properties of this fluctuation are derived on the Booker-Gordon assumption; namely, that the received wave is the sum of many little waves produced when the transmitter beam strikes "scatters" distributed in the troposphere. Expressions are obtained for the periods of the fluctuations in time, in space, and in frequency. These expressions extend closely related results obtained by Booker, Ratcliffe and others.

SCHIMPF, L. G., see R. L. Wallace, Jr.

¹ Bell Telephone Laboratories, Inc.

² American Telephone and Telegraph Company.

SMITH, K. D.¹

Properties of Junction Transistors, Tele-Tech, **12**, pp. 76-78, Jan., 1953.

TOWNSEND, J. R.⁵

What We Have Learned in 1952, Standardization, **24**, pp. 16-18, Jan., 1953.

TOWNSEND, J. R.⁵

What We Have Learned in 1952, A Report to the Joint Meeting of Standards Council and Board of Directors of ASA, A.S.T.M. Bull., No. 187, pp. 22-23, Jan., 1953.

VAN ROOSBROECK, W.¹

Large Current Amplifications in Filamentary Transistors, Letter to the Editor, J. Appl. Phys., **23**, pp. 1411-1412, Dec., 1952.

WALLACE, R. L., JR.,¹ L. G. SCHIMPF¹ AND E. DICKTEN¹

High-Frequency Transistor Tetrode, Electronics, **26**, pp. 112-113, Jan., 1953.

Sine-wave oscillators at frequencies up to 130 mc and tuned amplifiers with substantial gain at frequencies of 50 mc or higher are obtained by using junction transistors with an added connection to the base electrode biased negative at six volts.

¹ Bell Telephone Laboratories, Inc.

⁵ Sandia Corporation.

Contributors to this Issue

A. F. BENNETT, Western Electric Company, 1914-25; Bell Telephone Laboratories, 1925-. Mr. Bennett, Director of Station Apparatus Development since 1948, is in charge of the development, design, field trials, and studies of telephone instruments and sets, coin collectors, telephone booths, and station systems. During World War II, Mr. Bennett supervised the development and engineering work on a number of ordnance items. For this he was awarded a Presidential Certificate of Merit in 1946. He was a representative of the office of Scientific Research and Development and in that capacity served in the United Kingdom in 1943. He is a member of the A.I.E.E., the Acoustical Society of America, and the Physical Society.

R. W. BURNS, B.A., Indiana University, 1916; B.S. in E.E., Purdue University, 1918. U. S. Army, 1918; American Telephone and Telegraph Company, 1919-34; Bell Telephone Laboratories, 1934-. Mr. Burns has been engaged in the formulation of requirements for central office maintenance equipment, principally that used for testing exchange lines and trunks. During World War II he was in charge of a group preparing instruction books for teletype communication equipment for the armed forces. Professional Engineer, New York State. Member of Sigma XI and Eta Kappa Nu.

J. W. DEHN, E.E., Polytechnic Institute of Brooklyn, 1932. Western Electric Company, 1919-25; Bell Telephone Laboratories, 1925-. Switching Systems Development Engineer, 1952. Mr. Dehn has been principally concerned with the design and development of manual and dial telephone switching systems since joining the Bell System. During World War II he developed communication systems for the Signal Corps and trained military personnel in the operation and maintenance of this equipment. Since 1945 he has been engaged in No. 5 crossbar design. Professional Engineer, New York State.

R. F. MALLINA, M.E., Vienna Technical College, 1912. London Institute of City and Guilds, 1914. Wurlitzer Piano Company, Acoustical Engineer, 1925. RCA Victor Company, Head of Apparatus Development Department, 1929. Bell Telephone Laboratories, 1929-. At the

Laboratories Mr. Mallina worked initially in acoustical research where he developed the first magnetic telephone message recorder and the five-reed telephone set. From 1936 on he was engaged in fundamental development on machine switching apparatus, first on AMA, later with the wire spring relay project. In connection with the latter, he developed the solderless wrapped connection. Mr. Mallina is also a research associate at New York University, Department of Education.

W. P. MASON, B.S. in E.E., University of Kansas, 1921; M.A., Ph.D., Columbia, 1928. Bell Telephone Laboratories, 1921-. Dr. Mason has been engaged principally in investigating the properties and applications of piezoelectric crystals, in the study of ultrasonics, and in mechanics. Fellow of the American Physical Society, Acoustical Society of America and Institute of Radio Engineers and member of Sigma Xi and Tau Beta Pi.

J. W. McRAE, B.S. in E.E., University of British Columbia, 1933; M.S., California Institute of Technology, 1934; Ph.D., California Institute of Technology, 1937. Bell Telephone Laboratories, 1937-42, 1945-. U. S. Army 1942-45, where he attained the rank of Colonel and served as Deputy Director of the Engineering Division of the Signal Corps Engineering Laboratories. Returning to Bell Telephone Laboratories in 1945, Dr. McRae became Director of Radio Projects and Television Research in 1946; Director of Electronic and Television Research, 1947; Assistant Director of Apparatus Development I, 1949; Director of Apparatus Development I, 1949; Director of Transmission Development, 1949; Vice President, 1951. Legion of Merit, 1945. President of the Institute of Radio Engineers, 1953. Member of the A.I.E.E. and Sigma Xi.

T. F. OSMER, E.E, Polytechnic Institute of Brooklyn, 1935. Bell Telephone Laboratories, 1920-. As a member of the Physical Research Department until 1952, Mr. Osmer was primarily concerned with transducers, including transmitters, receivers, loudspeakers, and high quality microphones. During World War II he worked on military contracts, and since the war he has been occupied with carbon contact studies, and more recently with studies of the solderless wrapped connection, using photoelastic techniques.

R. C. PRIM, received a B.S.E.E. degree from the University of Texas in 1941, and M.A. and Ph.D. degrees from Princeton University in 1949. Following graduation from college, he was employed by General Electric Company in Schenectady until 1944, and then, as an ensign in the

Reserve, he joined the Naval Ordnance Laboratory at White Oak, Maryland. Here he conducted research on torpedo motion and control theory. He joined Bell Telephone Laboratories in 1949, and has conducted mathematical research on non-linear partial differential equations and served as a consultant on military projects. Dr. Prim is a member of the American Mathematical Society, the American Physical Society, Sigma Xi, and Tau Beta Pi.

E. D. REED, B.Sc., University of London, 1941; M.S., Columbia University, 1947. Ardente, Ltd., 1941-43; U. S. Army, 1944-46; Bell Telephone Laboratories, 1947-. Mr. Reed is engaged in the development and design of klystrons. Associate member of the Institute of Radio Engineers and member of Sigma Xi.

HARRY SUHL, B.Sc., University of Wales, 1943; Ph. D., Oriel College, University of Oxford, 1948. Admiralty Signal Establishment, 1943-46; Bell Telephone Laboratories, 1948-. Dr. Suhl conducted research on the properties of germanium until 1950, when he became concerned with electron dynamics and solid state physics research. Member of the American Institute of Physics and the American Physical Society.

R. H. VAN HORN, B.S. in E.E., Pennsylvania State College, 1937; M.A., Columbia University, 1947. Bell Telephone Laboratories, 1937-. Mr. Van Horn is a member of the Switching Apparatus Development Department and is in charge of the machine switching apparatus laboratory. He has previously been engaged in the development of underwater sound devices and the vibrating reed selector for mobile radio applications. Member of A.I.E.E. and Acoustical Society of America.

E. F. VAAGE, E.E., Technical University of Darmstadt, 1926; M.E.E., Brooklyn Polytechnic Institute, 1932; Royal Norwegian Air Force, 1918-19; Elektrisk Bureau, 1926-27; A. T. & T. Co., 1927-34; Bell Telephone Laboratories, 1934-. Mr. Vaage is a member of a group engaged in systems studies, an outgrowth of his previous work of evaluating transmission systems. Member of American Mathematical Society and A.I.E.E.

A. S. WINDELER, B.S., Rutgers University, 1930; Bell Telephone Laboratories, 1930-. Mr. Windeler has been engaged in the design and development of toll cable, including coaxial, video pair, and microwave, types. He is currently in charge of a group concerned with the development of expanded polyethylene insulated conductors for multipair cable.

1327
THE BELL SYSTEM
Technical Journal
VOTED TO THE SCIENTIFIC AND ENGINEERING
PECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXII

JULY 1953

NUMBER 4

THE L3 COAXIAL SYSTEM

Foreword

System Design

C. H. ELMENDORF, R. D. EHBAR, R. H. KLIE AND A. J. GROSSMAN 781

Equalization and Regulation

R. W. KETCHLEDGE AND T. R. FINCH 833

Amplifiers

L. H. MORRIS, G. H. LOVELL AND F. R. DICKINSON 879

Television Terminals

J. W. RIEKE AND R. S. GRAHAM 915

Quality Control Requirements

H. F. DODGE, B. J. KINSBURG AND M. K. KRUGER 943

Application of Quality Control Requirements in the Manufacture
of Components

R. F. GARRETT, T. L. TUFFNELL AND R. A. WADDELL 969

Abstracts of Bell System Technical Papers Not Published in this
Journal

1007

Contributors to this Issue

1015

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

S. BRACKEN, *President, Western Electric Company*

F. R. KAPPEL, *Vice President, American Telephone
and Telegraph Company*

M. J. KELLY, *President, Bell Telephone Laboratories*

EDITORIAL COMMITTEE

E. I. GREEN, *Chairman*

A. J. BUSCH

W. H. DOHERTY

G. D. EDWARDS

J. B. FISK

R. K. HONAMAN

F. R. LACK

J. W. MCRAE

W. H. NUNN

H. I. ROMNES

H. V. SCHMIDT

EDITORIAL STAFF

J. D. TEBO, *Editor*

M. E. STRIEBY, *Managing Editor*

R. L. SHEPHERD, *Production Editor*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; Alexander L. Stott, Treasurer. Subscriptions are accepted at \$3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXII

JULY 1953

NUMBER 4

Copyright, 1953, American Telephone and Telegraph Company

The L3 Coaxial System

Foreword

The articles in this issue are devoted to different phases of the development of a new system for the transmission and utilization of broader frequency bands on existing or new coaxial cables. This new system, which is called the L3 carrier system, represents the latest phase of development activities begun in the late twenties. It permits far more intensive exploitation of the cable medium than its predecessor, affording the option of providing, in each direction on a pair of coaxial tubes, either 1860 telephone channels or 600 telephone channels and a 4.2 megacycle broadcast television channel.

These results have been attained through wide extension of previous art. New electron tubes, transformers, inductors, and other circuit elements have been designed for extreme precision in respect to stability and other performance factors. Statistical quality control techniques are being applied to obtain the benefits of closely controlled distribution of the performance of circuit elements and system units. Fundamental to the program has been the devising of techniques for achieving hitherto unobtainable accuracies in the measurement of impedance, loss, phase and other transmission properties. To provide precise attenuation and delay characteristics over the wide frequency band, new techniques of network synthesis have been developed.

Refined system analysis and circuit design have derived maximum performance from component capabilities. The highest standards of

overall transmission performance and reliability have been adhered to. The new system is now in commercial service and large scale application is planned.

The following articles discuss (1) the over-all systems, together with its fundamental design problems, (2) the methods developed for equalization and regulation, (3) the broadband amplifying techniques, (4) the circuits for transmitting and receiving television, (5) the requirements established for controlling the performance of component elements, and (6) the application of these requirements in manufacturing.

E. I. GREEN

The L3 Coaxial System

System Design

By C. H. ELMENDORF, R. D. EHRBAR, R.H. KLIE, and
A. J. GROSSMAN

(Manuscript received March 31, 1953)

The L3 coaxial system is a new broadband facility for use with existing and new coaxial cables. It makes possible the transmission of 1,860 telephone channels or 600 telephone channels and a television channel in each direction on a pair of coaxial tubes. The principal system design problems and the methods used in their solution are described. The over-all system is described in terms of its components and their location in the system.

1.0 INTRODUCTION

The L3 coaxial carrier system is a new broadband transmission system capable of transmitting either 1,860 telephone message channels or 600 message channels and a 4.2-megacycle broadcast television channel, in each direction, on a pair of coaxials. The system is designed so that signals transmitted over any of these channels will meet high quality Bell System objectives after 4,000 miles of transmission.

The system is composed of auxiliary or line repeaters spaced at approximately four-mile intervals along the cable route and connecting terminal or dropping repeaters where telephone or television signals are translated to or from the L3 frequency band. Equalization equipment, power generating and power transmission equipment, and maintenance equipment are required at 100 to 200-mile intervals.

Planning and exploratory development for the system was started late in 1945 with the objective of designing a trunk route system which would provide the maximum channel capacity on the existing coaxial cable consistent with the state of the repeater art. At that time and for the next four years a large amount of new cable employing the 600 channel-three megacycle L1 coaxial carrier system was being installed or projected.¹

Since a major field of use of the L3 system was to replace the L1

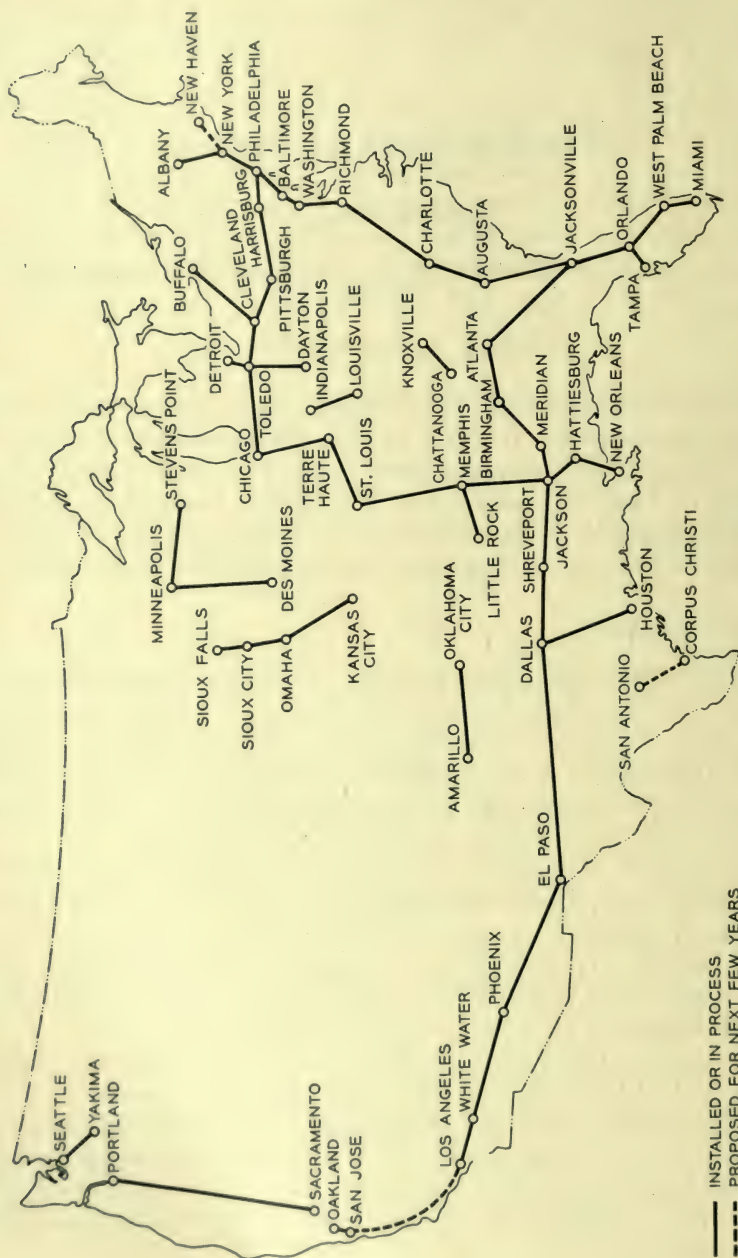


Fig. 1 — Coaxial cable routes.

system on existing routes, the design of the L1 system, the cable, and the cable route layouts presented the L3 system with a definite plant framework. The present day network of L1 coaxial systems is shown in Fig. 1. There are about 8,000 route miles of cable installed of which about 70 per cent consists of eight coaxials, the remainder consisting of six and four coaxials. About 70 per cent of this cable uses coaxials with a $\frac{3}{8}$ " diameter outer conductor, the present day standard. The remainder uses the older 0.27" diameter coaxials. All but a few miles of this cable is plowed into the ground or placed in underground conduit. A piece of a typical eight coaxial cable is shown in Fig. 2. Normally, the coaxials are included in a lead sheath with interstitial pairs which are used for

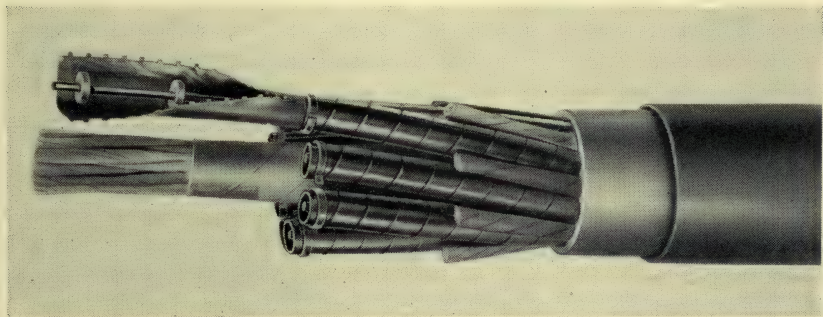


Fig. 2 — An 8-coaxial cable.

control purposes. In many cases additional quads are included in the cable for other types of transmission systems.

The broad objectives of the L3 system planning were:

1. The existing cable was to be reused. Thus, the cable loss and its variation with temperature, the cable irregularities due to manufacturing and splicing, and the power transmission capabilities of the cable became basic restrictions on the design of the L3 system.

2. The L1 telephone terminal equipment was to be reused. This equipment involves channel banks, group and super-group equipment and carrier supplies.² This limited the system planning to the use of frequency division multiplex on a single sideband carrier suppressed basis.

3. It would be desirable to reuse existing L1 repeater locations and buildings. The L1 auxiliary repeaters are spaced at eight mile intervals and housed in 6' x 9' concrete block huts. The L1 main repeaters are spaced at 40 to 160-mile intervals largely dictated by geographical and power transmission considerations.

4. Sufficient bandwidth should be provided so that a black and white television signal of at least four-megacycle quality could be transmitted simultaneously with 600-message signals, the message capacity of the L1 system. Alternatively, as many message channels as possible should be transmitted when there was no need for television service.

5. The channels should meet Bell System high quality signal-to-noise and equalization objectives after 4,000 miles of transmission.

Section 2 of the paper is devoted to a discussion of the principal system design problems and descriptions of methods used in solving these problems. Section 3 contains descriptions of the components of the system, their locations and their functions.

2.0 TRANSMISSION DESIGN

With a given cable loss, the line repeaters determine in large measure the bandwidth and quality of transmission and the economics of the system. The basic system plan therefore evolves from a consideration of the signal-to-noise and equalization performance — i.e., the transmission stability — that can be designed into the repeaters. This leads to the development of broad signal-to-noise and equalization analyses which guide and coordinate the system design.

2.1 SIGNAL-TO-NOISE DESIGN

Simply stated, the signal-to-noise problem is to adjust the repeater spacing and bandwidth of the system so that channel objectives can be met with the repeater noise, linearity, and gain performance that the electron tube and circuit art permit. In detail this means the following: (1) to translate the broad transmission objectives on message and television channels into detailed requirements on noise, specific modulation products and compression; (2) analyzing the amount of these interferences that result from various repeater design choices; (3) determining the effect of signal wave form and frequency allocations on both the channel requirements and the repeater performance; and (4) integrating these studies into a specific system design plan that meets the objectives.

2.11 *Telephone Channel Interference Objectives*

The amount of noise, tone interferences or crosstalk that is considered tolerable in telephone channels is generally determined by judgments involving the subjective reactions of representative observers to specific interferences on typical transmitted signals and by the cost of providing a given grade of service. The broad objectives for message

channels stem from early unpublished work on transmission standards. The interference and load capacity requirements for transmission systems involving large numbers of message channels were developed by Dixon, Holbrook and Bennett.^{3, 4} In effect, they provide techniques for translating channel objectives into linearity and power handling requirements on repeaters, taking into account the statistical properties of individual and multi-channel speech. Based on the data and techniques in these papers, the requirements on individual channels shown in Table I can be derived. These requirements in themselves form an important basis for the signal-to-noise design of the system. However, in a highly refined system design it is necessary to extend our notions of requirements somewhat further.

In the L3 signal-to-noise design the message channel requirements of Table I were used as the initial basis for study. However, when specific interferences of a complex nature were found to be limiting, the wave forms and the probability of their occurrence were examined in detail. As a result of these studies, two distinctive types of interferences were found to be important when the system is used to carry message and television signals simultaneously. The first of these, due to both second and third order modulation involving multifrequency key pulse signals and components of the television signal, has the characteristics of intermittent musical tones. The second, due to the second order difference products generated by the television signal components, produces tones in the message channels which vary in amplitude and frequency as the television signal changes with picture content. Both types of interference were generated in the laboratory and recorded on tapes. From these tapes, records were cut and then used in a series of subjective tests

TABLE I — SUMMARY OF MESSAGE CIRCUIT OBJECTIVES
(Allowable Zero Level Interference in 3 kc band)

Source of Interference	Type of Interference	dba (message Weighting)	dbm* Unweighted
Terminals.....	Largely spillover between channels, cross-modulation, and crosstalk	+32	-50†
Line.....	Noise and multichannel modulation	+36	-46†
Line.....	Unintelligible crosstalk and babble	+24	-58†
Line.....	Tones	+24	-61‡
Total.....	All sources	+38	-44†

* The translations from dba to dbm are effected by noting that a 3000 cycle band of flat noise with one milliwatt of power equals +82 dba and that one milliwatt of 1000 cycle single frequency is equal to +85 dba.

† Interference assumed evenly distributed over 3000 cycle band.

‡ Tones assumed to be at 1000 cycles.

which were made to determine the maximum permissible magnitudes consistent with other important message circuit objectives.

2.12 *Television Channel Interference Objectives*

The amount of noise and single-frequency interference that can be tolerated in a commercial grade television channel again depends on judgements involving the subjective reactions of observers and the cost of providing a given grade of service. The broad objectives are based on subjective measurements which have been reported on by Messrs. Mertz and Baldwin.^{5, 6, 7} From this work it has been determined that 95 per cent of the observers consider a signal-to-noise ratio of 40 db (composite signal to rms noise) tolerable, providing the noise has a frequency characteristic that rises about 11 db across the video band. Likewise the tolerable single frequency interference can be set at -70 db (peak sine wave below composite signal) if the interference falls below about one megacycle. The requirement becomes more lenient for interferences falling in the upper part of the band.

Again, for a refined system design, more detailed account must again be taken of the requirements on short duration interferences, the probability of interference occurring, and the exact frequency in the television spectrum that an interference occurs.

In the L3 signal-to-noise design the broad television channel objectives outlined above were used except when a specific complex interference was found to be limiting. For complex interferences, three additional types of requirement data were used; (1) tests were made to determine visual thresholds relative to steady tones of short bursts of energy such as occur in the television channel due to switchhook "bang-up" and multifrequency key pulsing signals in the message channels. Fig. 3 shows the relation between the steady state and transient requirement; (2) advantage was taken from the fact that interferences falling between the 15.75-kc line scan multiples of the television signal would be less interfering than unwanted energy falling directly at the line scan multiples; and (3) a judgement was used that the tolerability of an interference depends on its probability of occurrence. The judgement was not made on a quantitative basis but when an interference was found to exceed its requirement by a few db two or three times a day it was ignored in the signal-to-noise design.

2.13 *Frequency Allocations*

The final frequency allocations shown in Fig. 4 are a result of the signal-to-noise design. The principal features were determined on rather

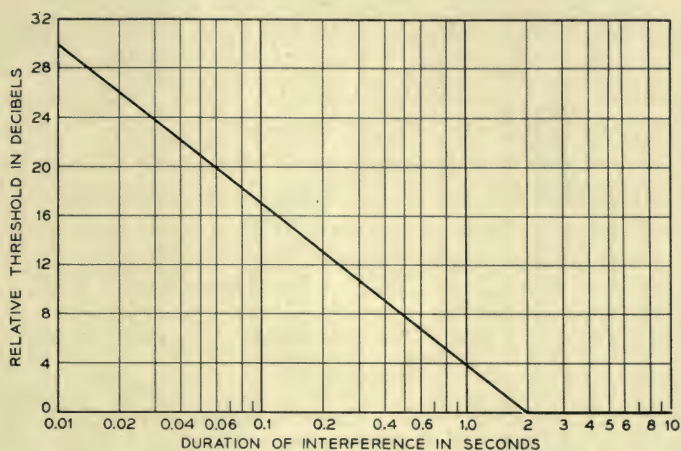


Fig. 3 — Television bar pattern threshold versus duration.

general grounds. When the system is arranged for combined television-message transmission, the television channel is placed above the message channels so that the second harmonic of the television carrier and its immediately adjacent side bands will fall at the top edge of the band where the requirement is more lenient. Likewise, the line repeater noise tends to rise with frequency as does the amount of noise that the television channel can tolerate. Details of the frequency allocations shown on Fig. 4 will be discussed in later sections.

Pilot frequencies, indicated on Fig. 4, are transmitted to control the transmission characteristic of the system as described in a companion paper.¹⁰ The frequencies, and the power at which the tones are transmitted, were selected on two bases; (1) where possible, frequencies used for similar purposes in the L1 system were selected for possible economies in pilot supply design and manufacture; these are the 556, 2,064 and 3,096-kc pilots; and (2) the transmission of these pilots should not materially degrade the signal-to-noise or load capacity performance of the system. The latter requirement led to a careful study of cross-modulation products involving the pilot frequencies to assure that message and television objectives would be met.

2.14 Repeater Performance

The details of the amplifier design and the factors which determine its performance are covered in a companion paper.⁸ For purposes of the signal-to-noise design it is sufficient to know the noise power vs frequency characteristic, the second and third order modulation coefficients

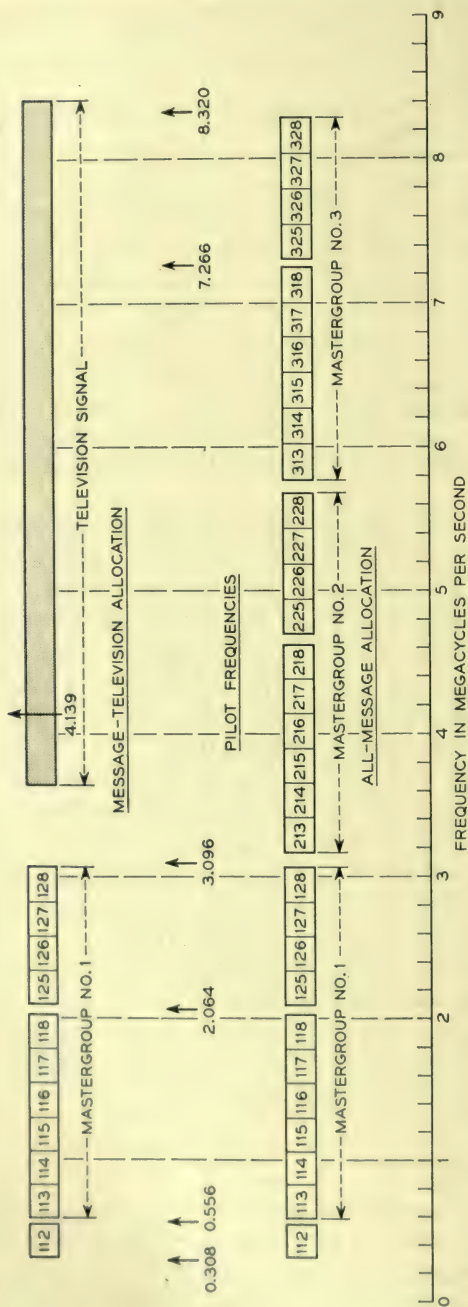


Fig. 4 — L3 frequency allocations.

of the repeater as functions of frequency and the overload performance of the repeaters. These factors depend on the repeater spacing and cable loss characteristic, electron tube parameters, achievable feedback, and the bandwidth to be transmitted. Thus, in the design procedures the dependence of these properties on repeater gain and bandwidth are determined and used in adjusting the system parameters for a final compatible design. Figs. 5 and 6 show the noise and linearity properties of the final L3 repeater. The four mile repeater spacing requires a repeater gain shown on Fig. 7.

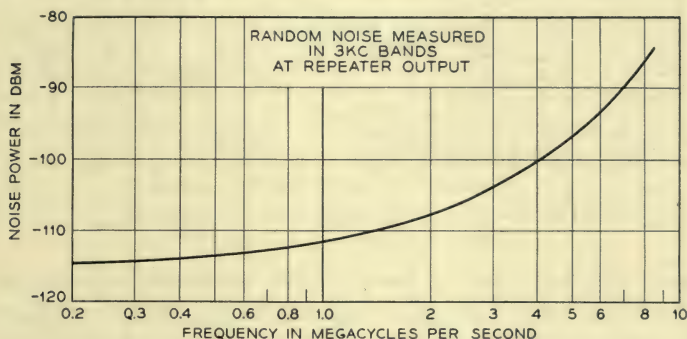


Fig. 5 — L3 Auxiliary repeater noise characteristic.

2.15 SIGNAL MECHANISMS

A signal-to-noise plan which contemplates transmitting the complex wave form of the combined telephone and television channels through 1,000 auxiliary amplifiers and about 200 flat amplifiers with performance factors that are variable with frequency will depend very strongly on the detailed analysis of the interactions between the signals and the repeater system characteristics. In developing this aspect of the signal-to-noise design four related phenomena had to be examined in detail.

2.151 *Intermodulation Between Signals in Different Parts of the Band*

In the classical multichannel modulation theory for a large number of message channels, the modulation noise generated by interaction of the speech signals due to the non linear characteristics of the amplifier is shown to be equivalent in interfering effect to random noise. In addition to this type of interference in message channels, cross modulation between components of the message and television signals result in a host of specific individual modulation products which have been examined by determining their amplitude, duration and probability of

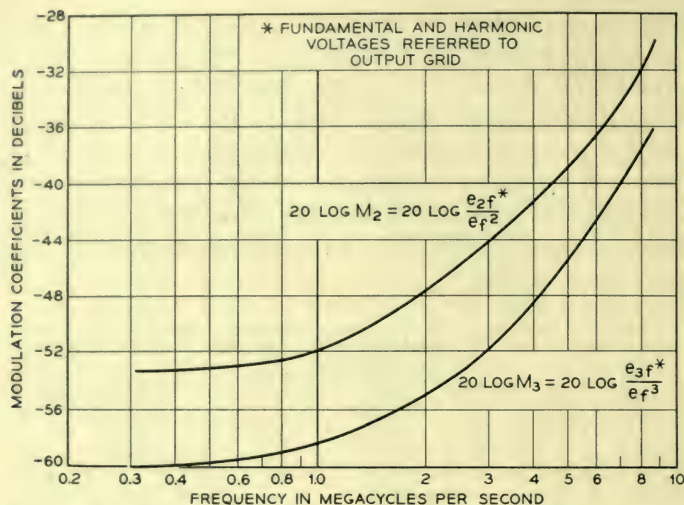


Fig. 6 — L3 amplifier modulation coefficients.

occurrence and then relating them to the requirements previously discussed. Approximately 400 different products or groups of products were studied in the design of the L3 system. All but about thirty of these were found to be of negligible importance for the signal levels and frequency allocations being given serious consideration. On final analysis six of these thirty products were found to be controlling in establishing system levels. Fig. 8 shows the generating signals and the products they form for the six most critical products. The exact way in which the critical pro-

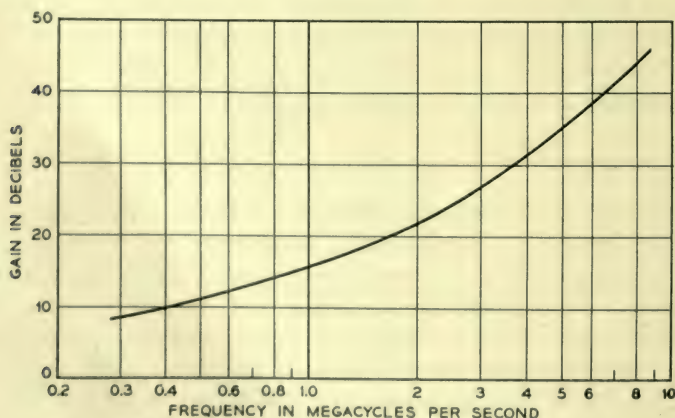


Fig. 7 — L3 repeater gain characteristic.

ducts entered into the determination of signal levels and frequency allocation will be discussed later.

2.152 Location of the Television Carrier Relative to the Telephone Channel Carriers

Among the important modulation product types is one formed by difference frequencies involving components of the telephone and television signals, see Fig. 8(d). These interferences fall back into the telephone band and are of different magnitudes depending, among other things, on which components of the television signal produce them; those

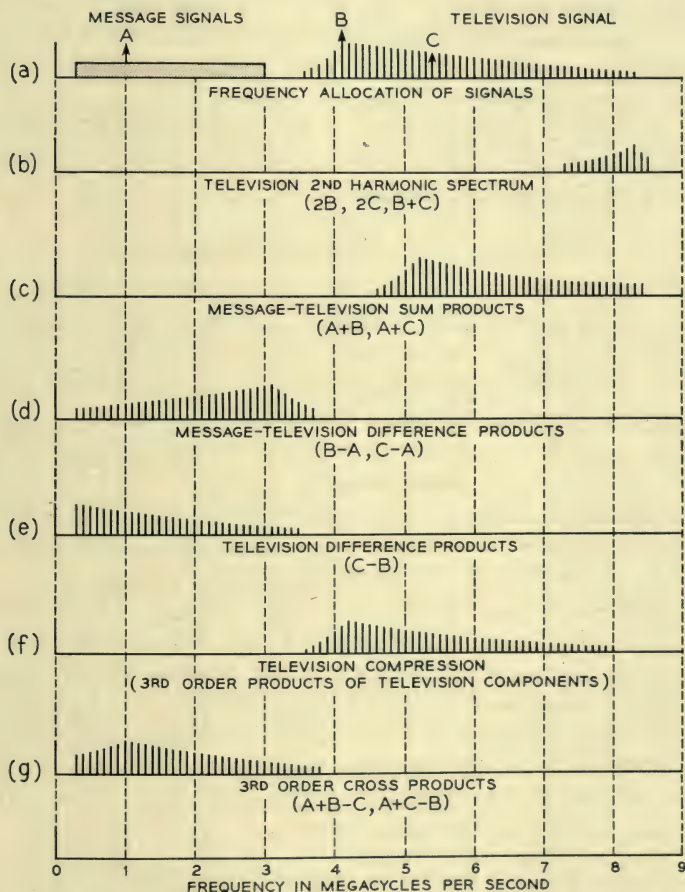


Fig. 8. — L3 coaxial system. Critical modulation products in combined message-television application.

produced by the television carrier and adjacent line scan multiples are by far the strongest.

The energy in a disturbing telephone channel tends to be concentrated near the 1,000-cycle point in the voice frequency band. By careful choice of the television carrier frequency, the difference products produced by cross modulation between telephone signals and the high magnitude television signal components can be made to fall at frequencies such that the high energy portions of these products are greatly attenuated by the cut-off characteristics of channel filters.

The message channels are spaced at 4-kc intervals controlled by carrier frequencies which are multiples of 4 kc. To obtain the maximum advantage from the channel filter cut-off characteristic as described above, it was found desirable to set the television carrier frequency 1 kc below a 4-kc multiple. A direct result of this allocation is a gain of 12 db in television signal-to-noise performance over what could be realized if the carrier had been set at a 4-kc multiple. Such an allocation would have required a 12 db lower magnitude of television signal in order to meet the message channel objectives.

2.153 *Addition of Modulation Products Along the Line*

It has been established by analysis and experiment, that in a multi-repeater system second order modulation products tend to accumulate on a power basis while certain third order products tend to add on a direct or voltage basis. This direct addition of third order products depends on the slope of the phase curve being the same over small frequency intervals from repeater to repeater. In multi-channel telephone systems, the locations of channels in the frequency band are shifted at intervals along the line to avoid this direct addition of third order products. In the combined telephone-television application of the L3 system the $A+B-C$ product illustrated in Fig. 8(g) is formed. Since the B and C components are television line scan multiples which cannot be shifted in location, certain components of this type product would add directly in a 4,000-mile system. If this were allowed to take place the requirements would be exceeded by many db. However, by placing the delay distortion equalization only in the television band at approximately 200-mile intervals the phase of these products can be shifted so that rms addition of products accumulated over several 200-mile links of the system may be assumed.

2.154 *Wave Form of the Transmitted Television Signal*

Early studies of L3 led to the conclusion that the most economical method of transmitting the television signal would be by amplitude

modulation of a carrier with one sideband partially suppressed, i.e., vestigial sideband transmission. There remained, however, three major problems for detailed study; (1) the transmission of dc components of the video signal; (2) the per cent modulation of the carrier which for convenience is defined in terms of "excess carrier ratio", the ratio of the peak (white) signal to the peak-to-peak composite signal as measured in the carrier frequency envelope; and (3) the sign or sense of modulation, that is, whether increasing or decreasing brightness should correspond to increasing signal voltage on the high frequency line. Typical waveforms illustrating the alternatives are shown in Fig. 9.

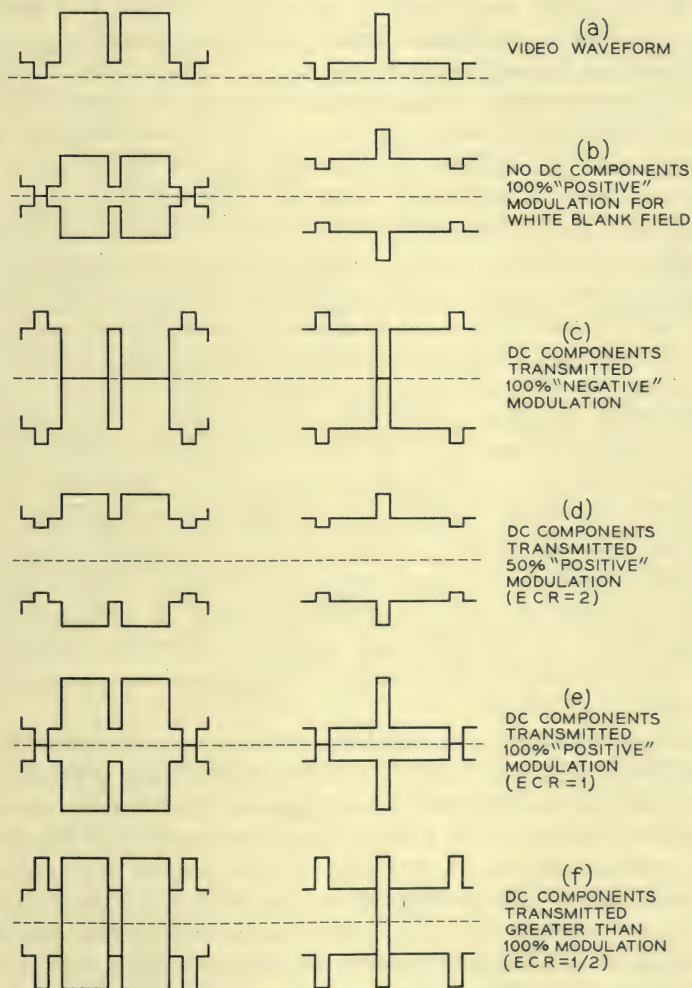


Fig. 9 — Typical television signals. Alternative carrier frequency waveforms.

The solution to each of these problems required an understanding of how the various alternatives would be affected by the system noise and linearity performance and an understanding of representative television viewing tube performance with respect to susceptibility to different types of interference. In analyzing the effect of system performance on these problems, it was found that non-linearity (cross modulation) would produce interferences in the television band which, while very complex electrically because of the effect of cross modulation involving line scan components of the signal, would produce the same effect on viewing tubes as single frequency interferences, i.e., bar patterns. Further simplifications were made in the analysis when it was found that such interferences were most visible in relatively large areas of television pictures having essentially constant brightness. During the time intervals corresponding to such areas, the video frequency voltage of the television signal is essentially constant and therefore, in the cases of interest, it could be assumed that the magnitude of the television carrier would also be constant during such intervals. Thus, to compute the magnitude of any modulation product which falls into the television band and which has as one of its components the television signal itself, it is found convenient to use in the computation the magnitude of the television carrier corresponding to either black or white portions of a picture signal. (The reason for intermediate shades of gray being less susceptible than either black or white is discussed below).

To evaluate the effect of television viewing tubes on wave-form problems, a number of tests were made to determine blank field threshold values of single frequency interference as a function of frequency for typical viewing tubes. Furthermore, judgements were made as to what might be expected of future viewing tubes with respect to achievable high light brightness, contrast ratio, and operating characteristics. As a result of these tests and judgements, a series of requirements were derived on the basis of long range objectives to be met for these projected characteristics. The results of these tests and judgements are summarized in Table II.

Using the parameters and methods of analysis outlined in the preceding paragraphs, the relative system performance achievable with each of the carrier frequency wave forms of Fig. 9 was computed or determined by observation. For example, these wave forms are all drawn to the same peak-to-peak amplitude. If we assume that the coaxial system is limited only by the peak amplitude transmitted we may use Fig. 9 to determine relative signal-to-noise performance directly by measuring the peak-to-peak magnitude of the composite signal voltage (sync tip to white) transmitted.

Fig. 9 may also be used to obtain relative modulation performance. For this purpose, the following factors must be considered; (1) the magnitude of the signal generating the interference ("black" or "white" carrier magnitude); (2) whether the interference is proportional directly or to the square of the carrier magnitude; (3) relative interference sensitivity in black or white portions of the picture; and (4) deviations from the Weber-Fechner law as the brightness is varied over its full range. The relationships among these factors were used to establish that for all cases of interest, bar patterns due to cross modulation are always more interfering in either black or white portions of a picture than in an intermediate gray area.

Table III shows the relative system performance for the five carrier frequency wave forms of Fig. 9. For comparison purposes, the signal-to-noise and signal-to-bar pattern ratios are all related to Fig. 9(f).

TABLE II — TELEVISION VIEWING TUBE CHARACTERISTICS ASSUMED FOR L3 SIGNAL-TO-NOISE ANALYSES

1. Brightness-grid voltage characteristic of viewing tubes follows 5/2 power law:
 $B \propto e_g^{5/2}$.
2. Maximum high light brightness of viewing tubes will be 150 foot lamberts.
3. Contrast ratio of viewing tubes will be 150:1.
4. Viewing tubes will have interference sensitivities which vary with brightness in accordance with the characteristic of Fig. 10.
5. The visibility of bar patterns will decrease with frequency in accordance with the characteristic of Fig. 11.
6. Deviations from the Weber-Fechner law may be assumed to follow the curve of Fig. 12. This law states that "the minimum change in stimulus necessary to produce a perceptible change in response is proportional to the stimulus already existing."

It is obvious from Table III that the signal is transmitted most efficiently at an excess carrier ratio of one half. The wave form of Fig. 9 (f), which illustrates excess carrier of one half, is the one used in L3. Television terminal circuit problems arising from this choice of carrier frequency wave form are discussed in another paper.⁹

2.16 *Signal Levels and Repeater Spacing*

In a broadband system like L3, the problem of determining the repeater spacing is made complex by the large number of parameters that must be considered. The approach to this problem that has been used to advantage in the L3 design is to assume several reasonable values of repeater spacing and determine for each the system performance achievable with various combinations of important parameter values. This method also permits evaluation of the effects on repeater spacing due to variations in parameters so that it is possible to form judgements as to the most economic design.

TABLE III — RELATIVE PERFORMANCE OF ALTERNATIVE TELEVISION WAVEFORMS

Waveform*	Relative Signal-to-Noise Ratio in db†	Relative Signal-to-Modulation Ratio (Bar Patterns) in DB‡	
		Group 1	Group 2
9b no dc	+10.2	+11	+11.3
9e neg. mod.	+6	+9.5	+12.5
9d ECR = 2	+12	+14	+15.5
9e ECR = 1	+6	0	+6
9f ECR = 1/2	0	0	0

* The waveforms are numbered to correspond to those given on Fig. 9.

† All values referred to E.C.R. = $\frac{1}{2}$; plus values indicate poorer performance.

‡ All values referred to E.C.R. = $\frac{1}{2}$; Group 1 products are those whose magnitudes are directly proportional to the carrier magnitude. Group 2 products are those whose magnitudes are proportional to the square of the carrier magnitude.

One of the important factors in setting repeater spacing is the magnitudes at which signals are transmitted in the system and the relation between these magnitudes and signal-to-noise and repeater overload performance. In the all telephone system (1,860 channels), the telephone levels (db with respect to the transmitting toll test board) were set to optimize signal-to-noise performance. To avoid penalizing the channels in the upper part of the band where random noise tends to be much higher than at low frequencies, the levels of the three mastergroups are staggered. At the output of any repeater in the high frequency line, the nominal level of mastergroup No. 1 is -21 db, that of mastergroup

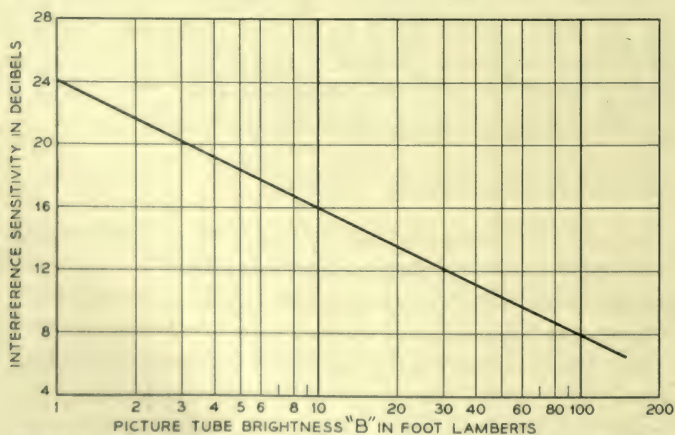


Fig. 10 — Picture tube interference sensitivity assumed for L3.

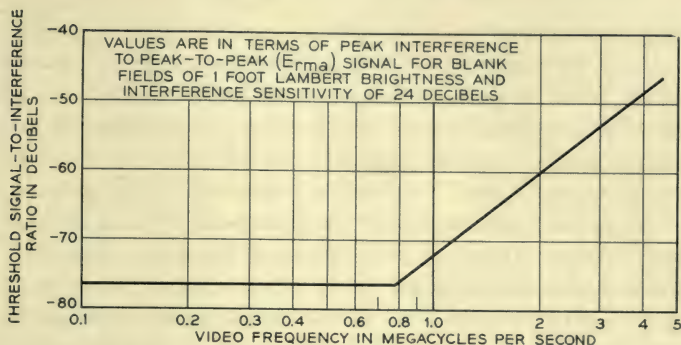


Fig. 11 — Threshold values for bar patterns.

No. 2 is -16 db and that of mastergroup No. 3 is -11 db. As a consequence of setting levels in this way, the random noise in the message channels is approximately 2 db higher on the average than modulation noise. It can be shown that with both second and third order modulation products contributing, and with third order somewhat predominant, this relation between random noise and modulation noise produces optimum signal-to-noise performance. With these levels, the 1,860 channel telephone system has approximately 6 db margin against repeater overload which, for L3 purposes, has been defined as the point at which the repeater modulation coefficients just depart from their constant small-signal values. The signal-to-noise objective of $+29$ dba at the -9 db level is met with about 2 db margin.

When the system is used to transmit television and message signals simultaneously, the level of the telephone channels in mastergroup No. 1 at the repeater output is the same as that of mastergroup No. 1 in the all-telephone application, -21 db. The most convenient measure of the television signal is the power of the unmodulated carrier at the output of a repeater. Its value is $+6$ dbm. Due to the inter-relations

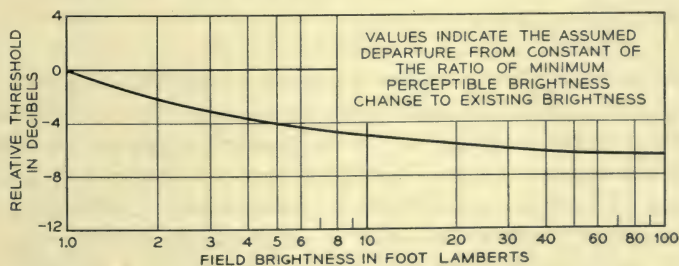


Fig. 12 — Assumed deviation from Weber-Fechner law.

between the two signals, the limitations on achievable maximum transmission levels or magnitudes arise from certain types of second order modulation products rather than from optimizing signal-to-noise performance as in the all-telephone application. One of these types consists of sum products of cross modulation between telephone and television signal components. These form bar patterns and, in so far as signal-to-interference ratio is concerned, are independent of the television signal magnitude. Thus, adjusting such products to equal the appropriate requirement has the effect of setting the maximum permissible magnitude or level of the telephone signal. The second type of limiting product is due to difference frequencies formed by cross modulation among the television signal components. These fall into the telephone channels and, after the telephone level has been set as described above, permit calculation of the maximum permissible television signal magnitude.

With signals set at -21 db level for telephone and $+6$ dbm unmodulated carrier for television, all of the critical products discussed in section 2.15 above and illustrated on Fig. 8 have adequate margin. The 40 db signal-to-noise objective for 4,000-mile television transmission is met with about 2 db margin and long haul (4,000 mile) message channels meet the $+29$ dba at the -9 db level objective with about 5 db margin. A margin of about 5 db is also realized with respect to repeater overload performance.

The single frequency pilots are adjusted to have the following values of power at the output of a transmitting amplifier:

7266 kc.	-16 dbm
8320 kc.	-26 dbm
All others.	-36 dbm

With these values, modulation products produced by cross modulation among the pilots and message and television signals all meet the appropriate objectives.

2.17 *Frogging of Message Circuits*

When signals, either message or television, are transmitted over long distances through many amplifiers in tandem, the accumulation of modulation products along the line becomes an important system problem for two reasons: (1) the accumulation of certain types of third order products tends to follow a direct or in-phase law and (2) the distribution of modulation products over the band produces more modulation noise in certain parts of the band than in others. Both of these cumulation problems are alleviated if, at intervals along the line, the signals

are shifted with respect to one another in the band, a process known as frogging.

In the L3 system, signal-to-noise performance is substantially improved by frogging the supergroups at intervals of about 800 miles. In the 1,860 channel all-message application, the busy hour signal-to-noise performance of 4,000-mile circuits is alike to within two db with all channels meeting the objective of +29 dba at the -9 db level. In contrast, if frogging were not specified, a substantial number of circuits (10 to 20 per cent would fail to meet the objectives while the performance of other channels would be better than required by six db or more.

When the system is used for combined message-television signals, the message circuits are frogged in supergroup blocks at approximate 800-mile intervals except for supergroups Nos. 113 and 114 which must be frogged at 400-mile intervals. This procedure is necessary to prevent second order sum and difference products of message and television signal components from cumulating excessively, especially those products which involve television signal components close to the television carrier. Frogging these supergroups more frequently than others results in a 3 db improvement in television signal-to-noise performance.

2.18 *Special Services Transmission*

During the early design stages, requirements based on the transmission of message and television signals were used to set repeater spacing, to determine the bandwidth and frequency allocations and to fix important design parameters of the amplifiers. Concurrently, the objectives for the transmission of telegraph, program, and telephotograph signals were studied and before the system design crystallized, analyses were made to assure that these special services objectives would be met.

In a few instances it was found that the special services objectives tended to dominate and the system requirements and design were adjusted accordingly. For the most part, however, channels which meet message circuit objectives are satisfactory for special services transmission. In L3, telegraph and telephotograph signals may be transmitted without restriction provided the proportion of these signals does not materially exceed the proportion now installed in the plant. Program signals may be transmitted in the 1,860 all-message arrangement without restriction but when television transmission is provided, program circuits are restricted to supergroups Nos. 113 and 114. This restriction is due to the fact that program circuits are usually more than 4 kc wide; interferences of high magnitude which normally fall between 3,300 and

4,000 cycles or below 300 cycles in message channels of supergroups other than Nos. 113 and 114 would fall close to 4,000 cycles in a program circuit where there is high susceptibility to interference.

2.19 *Uncertainties*

In the early stages of system design, firm decisions have to be made on such matters as repeater spacing, bandwidth, and component characteristics. These decisions must be based on a detailed signal-to-noise analysis which in turn involves many judgements of repeater performance parameters, tolerable system requirements and the effects of signal mechanisms on system performance. It would be easy and safe to engineer the system to provide enough signal-to-noise margin to cover the uncertainties in each of these judgements. Conservative engineering of this type could easily have justified a repeater spacing of three miles instead of the four miles actually chosen. Instead, an effort was made to estimate a "mid-range" or most probable value for each performance, requirement or mechanism factor entering into the signal-to-noise design. In addition, a "probable" uncertainty was estimated for each critical parameter. This was usually taken as one third of the maximum foreseeable error in the estimate. Finally, these uncertainties in electron tube modulation, realizable feedback, network impedances, channel requirements, interaction laws between signals and a myriad of other factors were all translated by the signal-to-noise analysis into their effect in db on the television channel signal-to-noise performance. On this basis, the "probable" uncertainties were summed on an rss basis to find the "probable" uncertainty in the overall design. Whereas the direct addition of the probable uncertainties gave a figure of about 20 db uncertainty in the design, the rss addition indicated about six db uncertainty. It was then argued that during the ensuing years of development the probability of finding all the judgements to be wrong in the same sense was extremely small. On the other hand it was deemed reasonable to provide enough margin so that there would be perhaps a 75 per cent chance of not exceeding the margin. Six db of margin was therefore provided, half by clear margin and half by having available economically feasible changes in system design such as a decrease in the telephone channel frogging interval. Any further error in judgement would then have to be taken up by degrading performance below the desired objectives. As the system design proceeded, the early judgements were changed in considerable measure. Likewise, numerous additional system parameters were introduced. However, at no point in the system plan-

ning was the balance of factors such that there was less than three db clear margin.

Margin handled in this way becomes a carefully husbanded asset of the whole system. In designing or analyzing a part of the system a major effort must be made to achieve the performance introduced into the initial determination of repeater spacing and bandwidth. The design of each individual part of the system cannot be allowed a margin which can be used up as the individual designer chooses.

2.20 *Equalization Design*

The term "equalization" is used to describe the process of obtaining flat gain and delay characteristics for the system transmission. The system and equipment designs to accomplish this function represent two of the major engineering features of the L3 system. In an overall sense, equalization includes the following: (1) determining deviation objectives for the gain and delay characteristics of the system and its component parts; (2) designing the auxiliary repeater so that the most economical over-all system equalization is obtained; and (3) specifying the location, form and control methods for the mop-up equalizers that are used at intervals along the system. Equalization and its related process, regulation, are the subject of a companion paper;¹⁰ therefore, in this paper only those aspects of equalization will be covered which are necessary for an appreciation of the over-all system design.

2.21 *Transmission Objectives*

The requirements on the gain characteristic of a band used for multi-channel telephony depend on two message channel objectives. One of these is that the gain of a message channel must not vary by more than two db over the 4-kc band. To meet this requirement, broad changes in the transmission characteristic of the message band are held to less than 0.5 db for 150-mile links. The second objective stems primarily from the need to transmit telephotograph signals. Since these signals are relatively intolerant of level changes, the transmission characteristics of working lines and protection lines are made alike to within ± 0.25 db.

The requirements on the gain and delay characteristics of the television band are based on the subjective determination that an echo delayed by about two microseconds or more in a representative picture is considered tolerable by 95 per cent of the viewers when the peak-to-peak voltage of the echo signal is 39 db below the peak-to-peak signal voltage.¹¹ The translation of this echo objective to allowable variations

in the gain and delay characteristics is straight forward if idealized sinusoidal deviations extending across the whole band are assumed.

In practice, the characteristics of the transmission deviations in a long repeater system are very complex and therefore, the idealized objectives are only a tentative guide in system design. Since we do not have a thoroughly satisfactory method of evaluating complex echo patterns, the exact nature of the final television mop-up arrangements will be determined after subjective tests on the interfering effects of echoes resulting from the complex transmission deviations of representative links of the system.

2.22 *The Mop-Up Plan*

The deviations from ideally flat gain and delay transmission characteristics may be classified in three broad categories; (1) fixed deviations; (2) slowly varying deviations; and (3) rapidly varying deviations. The distinction that is made between slow and rapid in the last two categories relates to the frequency of adjustment needed to meet system objectives. Those variations which require adjustment more often than once a week are considered rapid and those requiring adjustment at longer intervals are considered slow.

Corresponding to each of the three classifications of deviations is a set of equalizers, fixed, manually adjustable, or automatic under control of the pilot or a temperature sensitive element. Networks capable of fulfilling the functions of each are distributed along the line according to carefully prepared rules which enable system objectives to be economically met. The locations of these equalizers, their functions and general characteristics are illustrated in Fig. 13.

2.221 *Fixed Equalizers*

To the extent that the auxiliary repeater is designed so that its nominal gain compensates for the loss of four miles of coaxial, it may be considered as the first step of fixed equalization. In addition to the amplifier, the auxiliary repeaters are equipped with artificial lines, which are used to build out the loss of short sections to the equivalent of four miles of cable, and basic equalizers which provide for differences in the loss characteristics of different types of cable.

The second and final step of fixed gain equalization is known as a design deviation equalizer. Its function is to correct accumulated deviations due to the failure of the average auxiliary repeater to exactly

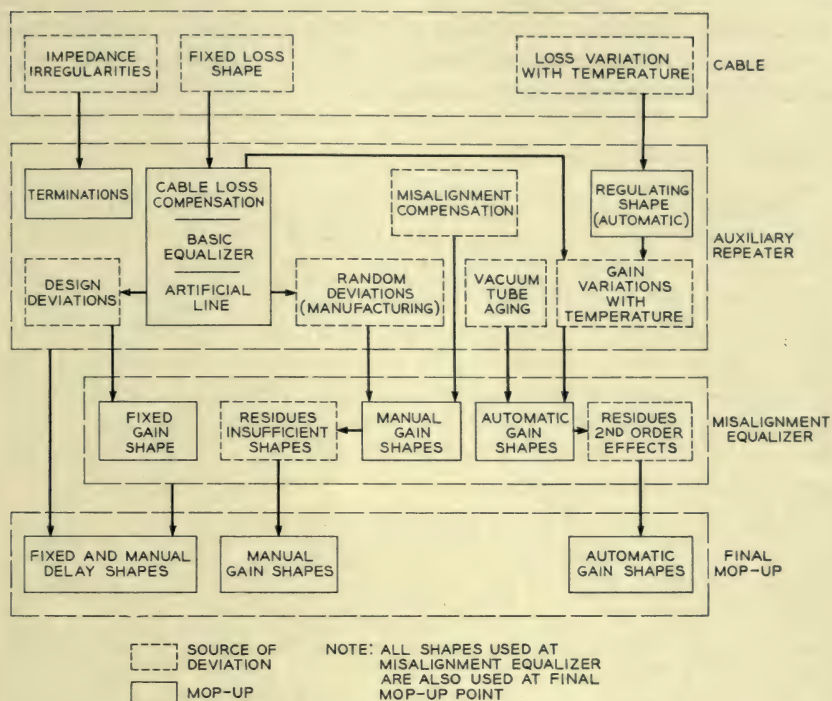


Fig. 13 — L3 coaxial system, equalization plan.

compensate for cable loss. These equalizers will be used at every mop-up point, at 40 to 120-mile intervals.

When television is transmitted, fixed delay equalizers are used at approximately 150-mile intervals. These equalizers compensate for the delay distortion introduced by the cutoffs of the auxiliary repeater sections.

2.222 Manually Adjustable Equalizers

The manually adjustable gain equalizers consist of networks whose loss-frequency characteristics are related to one another by a Fourier series type of representation. The number of terms of the series required to meet system objectives varies with different types of mop-up points and depends on the system application being provided for, all-message or combined message-television service. Equalizers of this type are used in mop-up points at 40 to 120-mile intervals along the line.

Manually adjustable delay equalizers are provided at approximately 150-mile intervals when television signals are transmitted. These equal-

izers supplement the fixed delay equalizers described above and are needed to trim the delay characteristic of the line in finer detail than would be possible with fixed equalizers.

2.223 *Automatic Equalizers*

The first step of automatic-gain equalization is provided at each auxiliary repeater. The nominal gain characteristic of the repeater is designed to match the loss characteristic of the coaxial at 55° F. The cable loss varies with changes in temperature; the variations, however, have a predictable characteristic, being very closely proportional in db to the square root of frequency. To compensate for these changes, the gain characteristic of the amplifier at auxiliary repeaters is adjustable and follows the loss of the cable under the control of a thermistor. Two types of circuits are used to control the current fed to the thermistor as described in a later section.

In the second step of automatic gain equalization, networks are provided to match system gain variations caused by electron tube aging and by changes in repeater hut temperatures. The loss characteristics of these equalizers are controlled by thermistors which in turn are controlled by the 308-kc and the 2064-kc pilots. These equalizers are used every 40 to 120 miles.

In the final step of automatic gain equalization, networks are provided to compensate for second order effects of the first three rapid variations described above, namely, cable loss variations, and repeater gain variations due to hut temperature changes and electron tube aging. The loss characteristics of these networks are under control of thermistors acted on by the 556-kc, 3096-kc and 8320-kc pilots. These equalizers are located at approximately 150-mile intervals.

The thermistors which control the loss-frequency characteristics of automatic equalizers are driven by regulators through a simple form of analog computer. The design and operation of this circuit is described in a companion paper.¹⁰

There is no automatic control of the delay characteristic in the system except that provided by the automatic gain equalizers. Every effort is made to have these equalizers match the transmission changes outside the band so that resulting delay changes in the band are minimized.

2.23 *Equalization System Considerations*

Whether the system is being equalized for telephone or television it is immediately apparent that the channel requirements described

earlier applied after 4,000 miles of transmission imply that, with no equalization, stability of the transmission characteristics of the individual repeaters would have to be of the order of a few ten thousandths of a db. Obviously, stabilities of this magnitude with changes due to temperature, electron tube aging and manufacturing processes cannot be achieved. Therefore, the equalization system design must be based on an economical balance between the cost of achieving repeater accuracy and stability and the cost of providing and maintaining an elaborate system of fixed, manual, and automatic equalizers.

The equalization problem involves so many variables that no attempt has been made to evolve a unified theoretical basis for evaluating the factors entering into this economic balance. However, in planning and designing the L3 system a number of principles and points of view have been developed which have guided the equalization planning.

2.231 *Misalignment*

The transmission objectives described above are determined on the basis of delivering satisfactorily equalized signals at terminals. In addition to this function the equalizers must limit the signal excursions along the line so that excessive noise or modulation is not accumulated in the repeater system. The amount of signal misalignment that can be allowed to accumulate before the first mop-up equalizer depends of course on the signal-to-noise allowance that has been made for this purpose. The amount of signal-to-noise performance allotted to misalignment must represent a balance between the reduced repeater spacing and increased complexity of equalizers that it costs and the increased spacing between mop-up equalizers and increased repeater deviations that it allows.

The engineering method for arriving at this balance represents an interesting example of system design by successive approximations. For example, the total gain area available (over an infinite frequency range) in a coupling network is inversely proportional to the capacity across the network and one of the important design choices is the extent to which one tries to utilize this area in the transmitted frequency band. The degree to which the available gain is concentrated in-band is called the resistance integral efficiency. In the very early stages of the amplifier design it was necessary to choose resistance integral efficiencies and frequency characteristics for the coupling networks. In a definite but complicated way these parameters are related to the sensitivity of the networks to element variations. Efficient networks give improved signal-to-noise performance but also increase the sensitivity to element

changes. By examining deviation curves for a number of specific network designs, tentative choices were made of 50 per cent resistance integral efficiency for the coupling networks and an allocation of about half the cable slope to the pair of coupling networks and the remainder to the feedback network. With an amplifier employing these networks a detailed study was made of the noise and modulation penalties at two frequencies resulting from misalignment in several lengths of line. This study indicated that with certain refinements in the repeater design the misalignment in twenty or more auxiliary repeater sections could be tolerated with a signal-to-noise penalty of about 2 db which was judged to be a reasonable allotment for this purpose. In addition, this study brought out: (1) that randomizing the variations of an element between its normal manufacturing limits resulted in a 4/1 reduction in the required misalignment allowance as compared with accepting large numbers of elements at one extreme of their limit; and (2) a small amount of gain adjustment at each repeater in the vicinity of the high magnitude television carrier would reduce the required misalignment allowance by about 2/1. Refinements on this plan for handling misalignment had to wait until the signal-to-noise and repeater design were crystallized. However, the study referred to above provided a powerful tool for evaluating proposed element deviations during the design period.

When the exact signal levels and the most limiting modulation products became known and when the repeater characteristics and final element deviations were determined it became possible to make a refined study of misalignment in terms of the noise and modulation impairment associated with specific signals and distortion products. At this point performance margins associated with specific interferences could be used to allow more or less misalignment of the particular signal components forming the interference. Likewise, amplifier deviations with specific frequency characteristics could be evaluated exactly in terms of their effect on the number of repeaters between mop-up equalizers. By studies of this type it was determined that the "A" or misalignment equalizers could be spaced at intervals not to exceed thirty-two auxiliary repeater sections.

2.232 *Distribution of Element Deviations*

The methods of statistical quality control used to monitor the process of manufacture provided the necessary techniques for obtaining the desired randomization of deviations. A companion paper¹² presents the techniques that were developed to apply the broad field of knowledge

on quality control to the specific needs of the L3 system. The most important point to appreciate in this connection is that the control of the process of manufacture (as well as the end electrical requirements) of individual elements is being used as a basic factor in the design of the system.

2.233 *Repeater Accuracy*

In developing the equalization plan it is a logical and straight forward operation to provide shapes and ranges in the equalizers that will compensate for the random variations of known elements. Likewise, real but indeterminate parasitic elements can be taken into account by specifying the final characteristics of the line amplifier feedback network and the equalizer fixed shapes (design deviation equalizers) on the basis of measurements on a rigidly controlled group of amplifiers that are deemed to be representative of the final product. However, having once specified the equalization on this basis the design elements and indeterminate parasitic elements must be held to the values and ranges upon which equalizer location, shapes and ranges are specified. This point of view has led to rigid mechanical control and the omission of component adjustments in the line amplifier which represent a departure from other transmission systems. These features are discussed in detail in the companion amplifier paper.⁸

2.3 NEW YORK-PHILADELPHIA TRIAL

The first installation of L3 has been made between New York and Philadelphia. Since the middle of 1952, this installation has been used to test components, to verify values of important system parameters used in system analyses, and to gather data for the further design and development of equalizers.

Random noise measurements have confirmed theoretical values (Fig. 5) to an accuracy of better than 2 db. In general, the measurements have indicated that the theoretical values have been conservative.

Measurements of system modulation performance, made with single frequency tones, also confirm the theoretical values used in analyses. Third order modulation measurements are in almost complete agreement with theory while second order measurements have been generally two to three db more favorable than the analytic values used.

Transmission measurements have confirmed that equalizer networks designed so far are satisfactory for systems to be installed in the near future. Further measurements are required to determine automatic

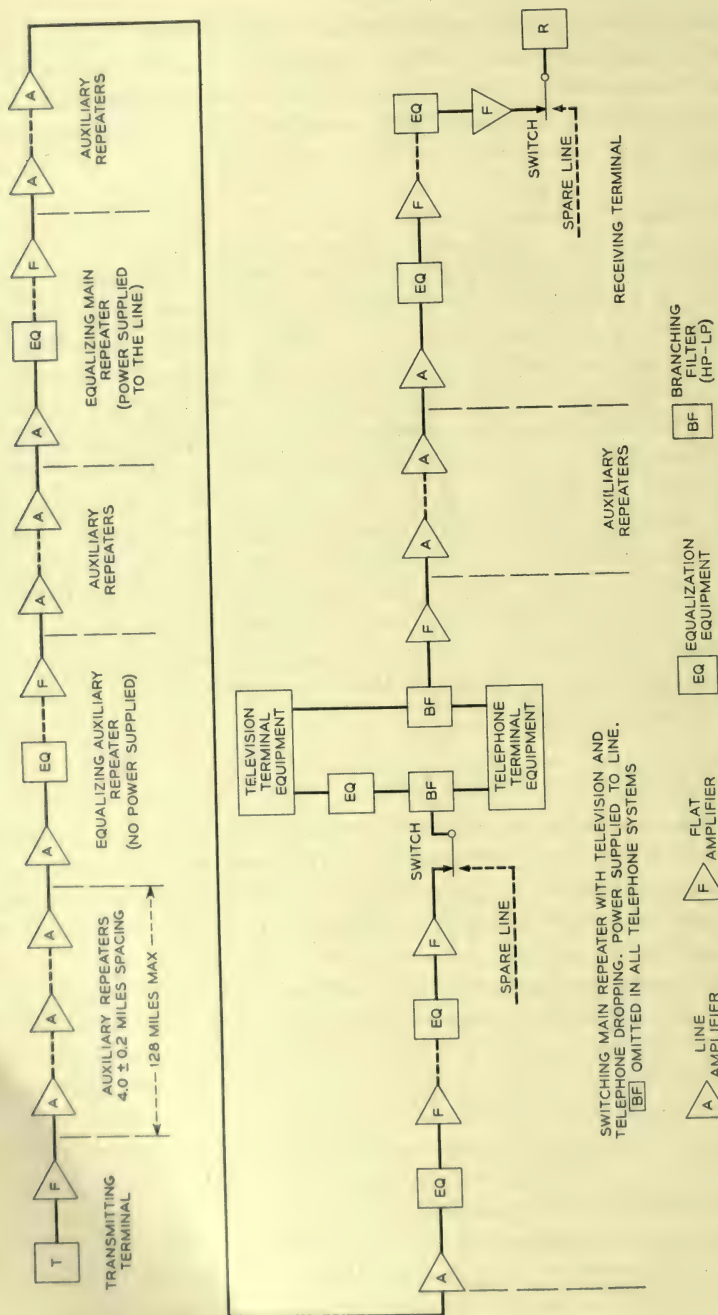


Fig. 14 — General layout of L3 repeaters.

equalizer shapes to correct for the second order effects of temperature changes and electron tube aging. Also under active study are the problems associated with final mop-up for long television systems.

3.0 SYSTEM DESCRIPTION

3.1 GENERAL

In the preceding sections on system design the functions of the auxiliary repeaters and the need for additional repeaters with varying amounts of equalization have been brought out. Fig. 14 shows the transmission layout of a typical L3 system. The auxiliary repeaters contain amplifiers and regulating equipment to compensate for the basic cable loss and its variation with temperature. Since such repeaters are dependent on the cable for their primary source of power they are called auxiliary repeaters.

At points in the system where additional first order equalization is required to reduce misalignment the complexity of the repeater equipment increases and such repeaters receiving power over the cable are called equalizing auxiliary repeaters.

The distance which power may be transmitted over the cable to the auxiliary repeaters is limited; therefore, repeaters at specified intervals must be capable of supplying power to the cable. These are called main repeaters. They may be equalizing main repeaters where only first-order equalization is required or switching main repeaters where lines are switched or circuits dropped.

3.2 AUXILIARY REPEATER

3.21 *Transmission Circuit*

The auxiliary repeater is the basic unit of the system and its design determines to a great extent the performance and economics of the system. A block diagram of such a repeater for transmission in two directions on two coaxials is shown in Fig. 15. The power separation filter (PSF) is a six terminal high pass-low pass filter designed to separate the high frequency transmission signals on the coaxial from the low frequency current transmitted on the center conductor to furnish primary power to the repeater power equipment. At the input to the repeater the low-frequency current is diverted to a power supply while the high-frequency current follows a path through passive networks to the input of the amplifier. At the output, the signal from the amplifier and the low-frequency current from the power supply are recombined in the power separation filter for transmission to the next repeater.

The power separation filters are basically simple designs, but the realization of the theoretical design was complicated by the following: (1) the components in the low frequency section must pass currents of about 1.5 amperes without change in characteristics, and must withstand potentials as high as 2,000 volts rms without generating corona noise; and (2) the components in the high frequency section must be such that the loss over the transmission band (300–8,350 kc) is small and stable and of such a shape that it is easily equalized. To meet these requirements stable inductors and capacitors with a minimum of parasitic resonances in the band were designed.

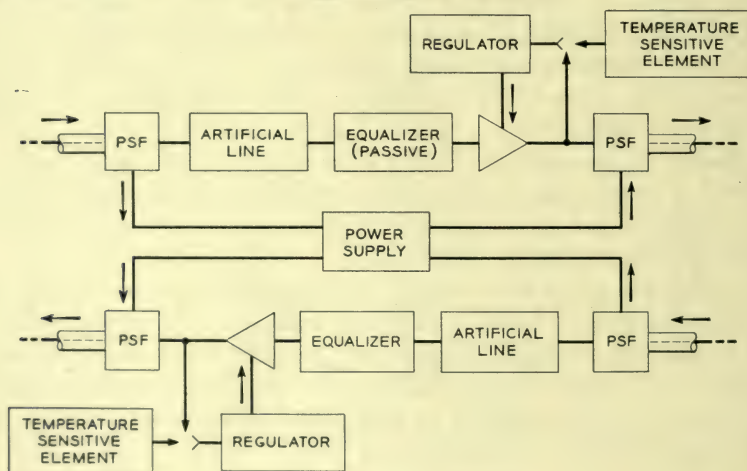


Fig. 15 — Auxiliary repeater.

The artificial line shown preceding the input to the amplifier is a passive network to build out the loss-frequency characteristic of a short cable section to be equivalent to the loss-frequency characteristic of 4.0 ± 0.2 miles of 0.375" cable or 2.87 ± 0.15 miles of 0.27" cable. These lines are provided in several different sizes, so that, where it is impossible physically to locate the repeater within the specified accuracy of 0.4 mile this accuracy can be obtained electrically. The design of the network is such that an accurate and stable characteristic is obtained with a minimum number of components.

The equalizer is a means for compensating for small variations in the transmission characteristics of coaxial cables due to variations in the physical construction of the cable. In the case of the most generally used cable, this equalizer inserts only a small flat loss.

The amplifier is of the feedback type whose gain frequency char-

acteristic is closely equivalent to 4.0 miles of 0.375" coaxial cable plus the loss of the other passive elements in the repeater. This unit is demountable without tools for maintenance and is sealed in a die cast housing as protection against moisture and dust. The detailed electrical and mechanical design are covered in a companion paper.⁸

The regulator may be one of two types. The first, called the auxiliary regulator, adjusts the gain-frequency shape of the amplifier in accordance with the magnitude of the 7,266-kc pilot transmitted along the line. The second type, the thermometer regulator, adjusts the gain-frequency shape of the amplifier under control of an element representing an average value of cable temperature. This element is a thermistor buried in the ground near the cable. Such a control is, obviously, not as accurate as pilot controlled regulation, but it is adequate for use at one-half of the auxiliary repeaters and its simplicity results in considerable saving in first cost and power requirements. The regulators are demountable units similar to the amplifiers. Their detailed electrical and mechanical design are covered in a companion paper.¹⁰

The pilot alarm unit is provided with auxiliary regulators to indicate pilot deviations beyond a predetermined limit. Its operation will be described a little later in connection with the discussion of alarm and control arrangements for the entire system.

3.22 *Power Supply*

Primary ac power for the auxiliary repeater is supplied on a constant current basis from the main repeater over the center conductors of the two associated coaxials. Power generating and control equipment used at the main repeater will be discussed in Section 3.6. At the auxiliary repeater, power supply equipment is required to convert the primary power to suitable voltages for heater, plate and bias use as shown on Fig. 16. Half of the input to the power supply is taken from each center conductor and the output of the power supply is used to power the entire two-way repeater.

The heater voltages are obtained by simple transformation which is complicated only by the fact that accurate and low loss transformers are required and the primaries of these transformers must withstand high ac potentials without generating corona noise which might be transmitted through the power separation filter to the input of the amplifier. Two separate transformers are used to split the load between the two center conductors even though the secondaries are connected together to feed the repeater. This arrangement eliminates one crosstalk path

at low frequencies where it is difficult and expensive to design power separation filters to meet the system requirements.

The dc plate and bias supply voltages for amplifiers and regulators could be obtained by conventional rectifier circuits except for one complication which such arrangements introduce. This complication is the fact that a rectifier terminated in a low-pass filter (conventional ripple filter) reflects a highly distorted current wave into the primary circuit. If the primary current is so distorted the various power supplies in the series circuit will be fed with other than a sine wave of current and will supply different voltages depending on the wave form. Since the heater power depends on the rms value of current while the dc output depends on the peak value of voltage, it is easily seen that the relationship between these two will change with the wave form of the applied current. Furthermore, the line loading to be discussed later must be calculated on the basis of a pure sine wave; appreciable harmonics in the line current tend to make it impossible to predetermine the loading to any reasonable degree of accuracy.

It was found that these problems could be avoided and the power factor of the power supply made very nearly unity if the rectifier (RECT 1) was terminated in a constant resistance load rather than a low-pass filter. This was provided by paralleling the conventional low-pass filter with a high-pass section terminated in the proper resistance load. To avoid wasting the power in this load a second rectifier was added (RECT 2). The dc output of this circuit is used in series with the main dc supply to provide the higher voltage required for the output stage of the amplifier. This rectifier must also be terminated resistively although its effect on the main current wave is less than that of the first rectifier, and the power dissipated is smaller. Since there was a further use for a small amount of power for bias in the regulators, RECT 3 was added to produce a regulated voltage in conjunction with a conventional gas tube circuit. This rectifier and its load provide the termination for the high pass section of the filter circuit for RECT 2. A second gas tube circuit is used to obtain a regulated bias supply for amplifiers and regulators from the 315-volt source. The loads on both gas tube circuits are fixed so regulation for variation in input voltage only is required. For this reason a low current, highly stable gas tube could be used.

3.23 *Power Loading*

The power transmission circuit of a power loop is essentially a resistance-capacity network at the power frequency. The line and the

power supply resistance are the resistance component and the line and power separation filter capacity to ground make up the shunt capacity. If the circuit were used in this form the primary current at each repeater would be different since it would be the vector sum of the current in the succeeding section and the current in the shunt capacitance. This is undesirable as the objective is to make all power supplies alike. A familiar solution is applied to this problem by inserting an inductive reactance in series with the line. A value of this reactance is chosen for each repeater to compensate for the current through the effective shunt capacitance and thus make the currents through each of the power supplies as nearly alike as possible.

To simplify the loading adjustment in the L3 system a continuously variable loading inductor was developed. This arrangement allows more accurate adjustment of loading without the complications of changing wiring taps in a high voltage circuit. The design of such an inductor presented formidable obstacles as a large range of variation was desired (20–120 mh), and relatively high currents and voltages were involved. The device used consists of the two inductors which may be rotated with respect to each other, so that the coupling between their magnetic circuits varies ideally between zero and 100 per cent. One inductor is inserted in each side of the power circuit and a net result is obtained which is equivalent to varying each inductor.

3.24 *Physical Description*

The type of auxiliary repeater generally used is shown in Fig. 17. It consists of a 6-foot cable duct framework upon which the component panels are mounted. It is completely wired in the factory. The lower third of the bay contains the power supply equipment while the upper part contains two transmission panels. One panel is provided for each direction of transmission and all of the transmission components of the circuit are found on these panels. The demountable units, amplifier, regulator, and pilot alarm unit are interconnected with plugs and jacks, so that they may be removed for maintenance. The other units are interconnected by screw-type terminals and cable as it is expected that they seldom will require maintenance.

Other types of repeaters will be available to meet special conditions such as manholes where sealed apparatus cases will be required to prevent damage due to water submersion, or telephone offices where standard 11'-6" frameworks are usually desired.

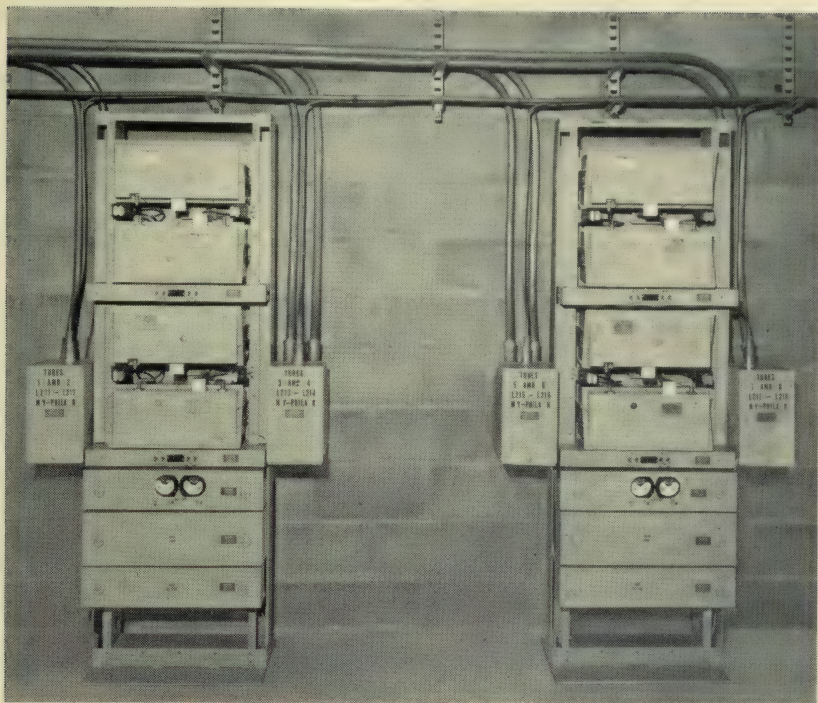


Fig. 17 — Typical auxiliary repeater in concrete block nut.

3.3 EQUALIZING AND SWITCHING REPEATERS

3.31 *Components*

The equalizing auxiliary and main repeaters use the same general types of transmission equipment as the auxiliary repeater except for the equalizers. They differ principally in the quantity of equalization equipment provided and the bay arrangements. Table IV lists a summary of the basic transmission units in each repeater.

At these repeaters a line amplifier is used as a receiving amplifier to compensate for the previous section of cable. Flat amplifiers are used as transmitting amplifiers and to compensate for the loss introduced by the equalizers. They have a flat gain-frequency characteristic and no provision for pilot control of their gain. Their design is very similar to that of the line amplifier and is covered in the companion amplifier paper.⁸

TABLE IV—SUMMARY OF HIGH-FREQUENCY LINE EQUIPMENT
AT REPEATER STATIONS

	Auxiliary	Equal- izing Aux.	Auxil- iary Main	Switching Main (Tele- phone only)	Switching Main (Tele- phone — TV
Line amplifier.....	1	1	1	1	1
Flat amplifier.....	0	2	2	5	5
Auxiliary or thermometer regulator.....	1				
Manual equalization (Number of terms) ..		10	10	15	25
Automatic equalizers	1	3	3	6	6
Regulators for equalization (kc).....		7266	7266	7266	7266
		308	308	308	308
		2064	2064	2064	2064
				3096	3096
				556	556
				8320	8320
Design deviation equalizer.....		1	1	1	1

The functions of the fixed, manual and pilot controlled equalizers are noted in Section 2.22 of this paper and discussed in detail in the companion equalization paper.¹⁰

3.32 Equalizing Auxiliary Repeaters

This type of repeater will be found after a maximum of thirty-two auxiliary repeaters provided power feed to the cable, dropping, or switching is not required (Refer to Fig. 14). The major components provided are covered in Table IV. In addition to these items, power separation filters, basic equalizers and artificial lines identical with those in auxiliary repeaters are used. A pilot alarm unit is also included to monitor each of the three regulators and transmit an alarm when any one of the controlling pilots has deviated beyond a given limit.

Power for these repeaters is obtained from the cable just as in the case of the auxiliary repeater and much of the same type of equipment is used. However, due to the larger amount of power required and the layout of the repeater the auxiliary repeater power units have been re-packaged to provide the optimum arrangements for leads carrying high current or critical bias supply circuits.

The design of panels used in this repeater was dictated by the general scheme conceived for the switching main repeater where the maximum amount of equipment is required. This arrangement involves the use of both sides of a duct-type frame. A single panel (again called the transmission panel) is used, but an amplifier is mounted on one side and a

regulator is mounted on the other. This requires access to both sides of the bay, but results in an overall saving in the number of bays and overall floor space.

All of the transmission components and a heater and bias supply unit are mounted in one 7' bay for each coaxial. The plate and primary ac power for two of these bays is mounted in another 7' bay.

3.33 *Equalizing Main Repeater*

This repeater contains exactly the same transmission equipment as the equalizing auxiliary repeater (see Table IV). It differs in the function noted before, that is, it is equipped to feed power to the cable. The equipment to perform this function will be described later in the paper. Since the repeater can feed power to the cable it can also supply the power for its own operation. This power is derived from the primary ac supply used for the line by means of conventional metallic rectifiers for dc circuits and transformers for the ac heater supplies. These power supplies are not a part of the power loop containing auxiliary repeaters, so no special arrangements are required to obtain good waveform or high power factor.

The equipment arrangement uses the same units as the equalizing auxiliary repeater, but here conventional 11'-6" frames are used.

3.34 *Switching Main Repeater*

Usually, this type of repeater is supplied at the point where circuits are dropped or terminated. In order to permit switching from a working line to a spare line in case of trouble, see Section 3.4, more complex equalization is required so that the lines will be as nearly alike as possible when the switch takes place. Furthermore, the signal delivered to the terminal must meet equalization limits that will result in a satisfactory grade of service.

Where the repeater is part of a system required to transmit only message circuits the basic equipment shown in Table IV (telephone only case) is required. In addition to these units facilities are provided which indicate and alarm pilot levels and provide for patching and other maintenance arrangements. Since this repeater always feeds power to the cable it uses the same power arrangements as the equalizing main repeater.

When the system is being used for the combined telephone and television signal this repeater is the same as the "all telephone system" repeater except that it has additional equalization equipment to adjust

the system for the more stringent television requirements. Furthermore, line connecting equipment consisting of branching filters and additional equalization is required. Branching filters are used to separate the telephone and television bands so that they can be transmitted to their respective terminal equipment. These are combined high-pass low-pass structures complicated by strict requirements on stability of the gain-frequency and delay-frequency characteristics. Delay equalization for the line sections must also be provided in the television branch and a large part of this is combined with the branching filters. Other components required for long television systems are adjustable gain and delay equalizers and associated amplification.

The same type of equipment is used as that described for the other repeaters except that a number of additional transmission panels are required to mount the additional amplifiers and regulators associated with the equalizers. Two 11'-6" bays are used to contain the equipment for one through coaxial. One bay contains the receiving equipment which precedes the line switch. The other bay contains the transmitting equipment (transmitting amplifier and hybrid) and any line connecting equipment for combined systems. Fig. 18 shows a typical main repeater installation.

3.4 AUTOMATIC SWITCHING*

In order to preserve transmission in the event of the failure of a component of the system and for transmission maintenance purposes, one coaxial in each direction is operated as a standby. An automatic switching system is provided to permit substitution of the standby line for any of the working lines. The lines are switched at the input to the transmitting amplifier and at the output of the receiving amplifiers and equalizers. (See Fig. 14).

At the receiving end of a switching section, equipment is provided whose function is to recognize failure of a working line and initiate the switching circuits. Information as to the transmission conditions of the system exists in the pilot regulators, the output of which controls a sensitive relay with high and low limit contacts. The operation of one of these relays provides the switching system with the information that transmission has failed or been seriously impaired. It is necessary to make a switch as rapidly as possible in the case of a total failure in order to reduce the effect upon the transmission circuits. As the relays take appreciable time to operate, the receiving switch equipment is designed

* Material written by P. T. Sproul.



Fig. 18 — Typical main repeater installation.

to operate directly from the dc output of the regulator associated with the 7,266-kc line pilot. This permits complete switch operation in about 15 milliseconds.

Upon receipt of information from the regulators that one or more of the pilots have gone out of limits, the switch initiator signals the transmitting switch control equipment at the transmitting end of the switching section as to which line has failed. Signalling is accomplished by the use of tones in the 280 to 296-kc range which are transmitted over all coaxials in parallel in the reverse direction. Use of the coaxials for transmission of these signals obtains maximum speed of operation of the switch. All paths are used in parallel to preclude failure of the switch if one channel in the opposite direction would be inoperative.

The transmitting switch control equipment causes the transmitting end of the standby line to be switched in parallel with the line in trouble and then signals the switch initiator that this has been done. This verifier tone actuates the line switch at the receiving end to complete the switch.

When the trouble clears, the switch is released as the initiator checks every minute to see if service on the working line can be restored. In the event of a prolonged trouble, the switch can be locked manually and the initiator will no longer attempt to restore to normal. Release of the switch is accomplished by the transmission of a release tone to the transmitting end while a checking tone returned by the transmitting end indicates completion of release and readies the switch initiator for further switching.

For maintenance purposes, manual operation of the switching equipment is provided. In effect, manual switches are made by simulating a failure. Alarm features are provided to indicate to the operating personnel failure of the coaxial system or failure of the switching equipment. Care has been taken in the design to insure that failure of the switching equipment in no way affects transmission except by removing the protection afforded by the presence of the switching facility.

One 11'-6" bay is required for the switch control equipment for each direction of an 8-coaxial system. The line switches are mounted in the miscellaneous bay of the main repeater lineup.

3.5 TERMINALS*

Television terminal equipment, which is required to modulate the video frequency signals to and from the high-frequency line, is described in a companion paper⁹ and therefore, will not be discussed here.

* Material written by C. G. Arnold.

The telephone terminals consist of modulators (and related transmission equipment) and carrier and pilot generating equipment. The transmission components of a terminal for an all message system are shown on Fig. 19. The channel, group and supergroup equipment are designs previously used in the L1 system. The designs of the submastergroup and mastergroup units employ circuit arrangements similar to those used in the supergroup equipment. The greater bandwidths, higher frequencies and more severe stability requirements required new components and improved circuit and layout techniques.

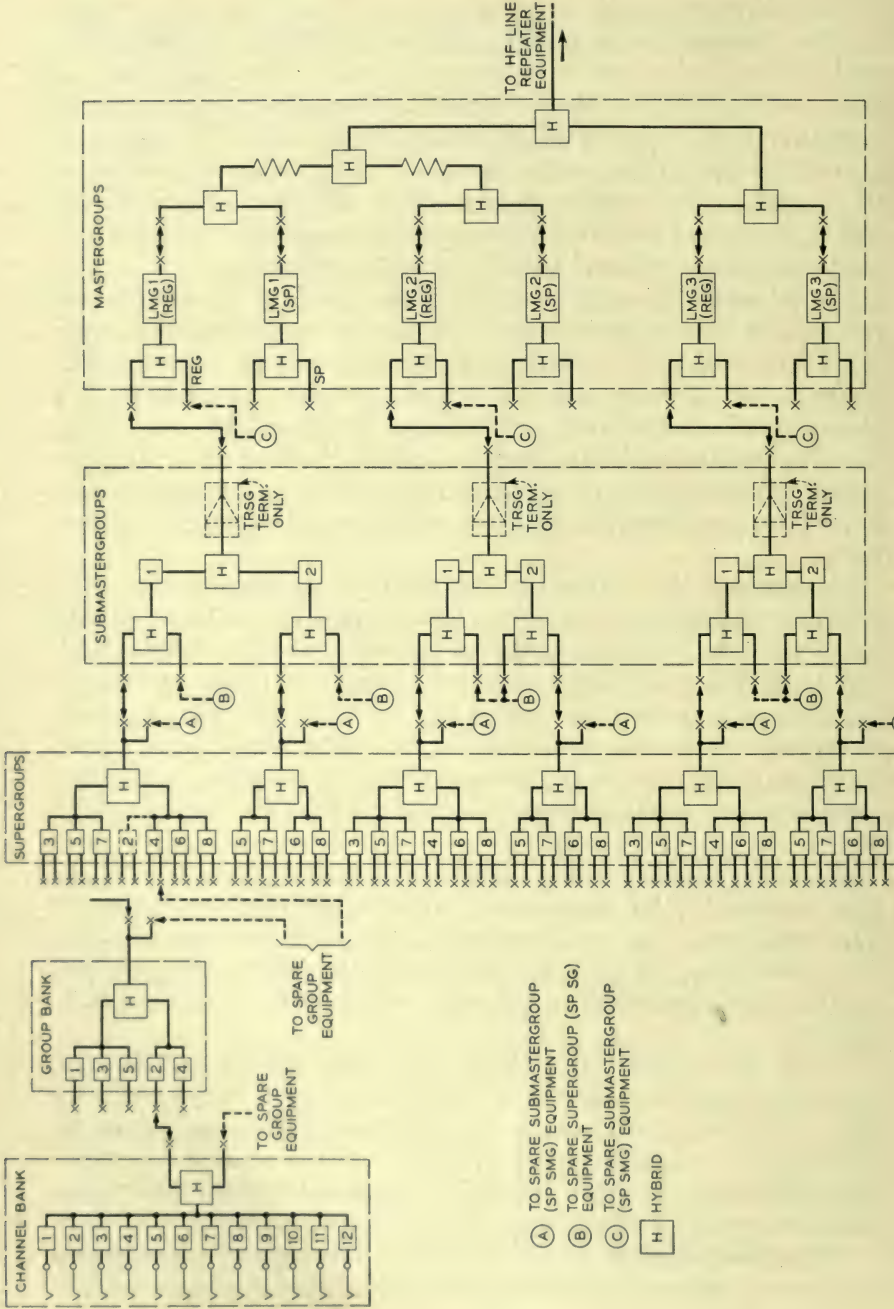
Fig. 20 shows the modulation steps and location in the frequency spectrum of the supergroups, submastergroups and mastergroups when the L3 system is used for telephone and television or all telephone. Mastergroup one comprises the first ten sixty-channel supergroups. This mastergroup is placed directly on the line in the 564 to 3,084-kc frequency band for both the telephone-television and all-telephone cases. When the system is used entirely for telephone, two additional mastergroups are formed by modulating mastergroup one up into the desired frequency bands.

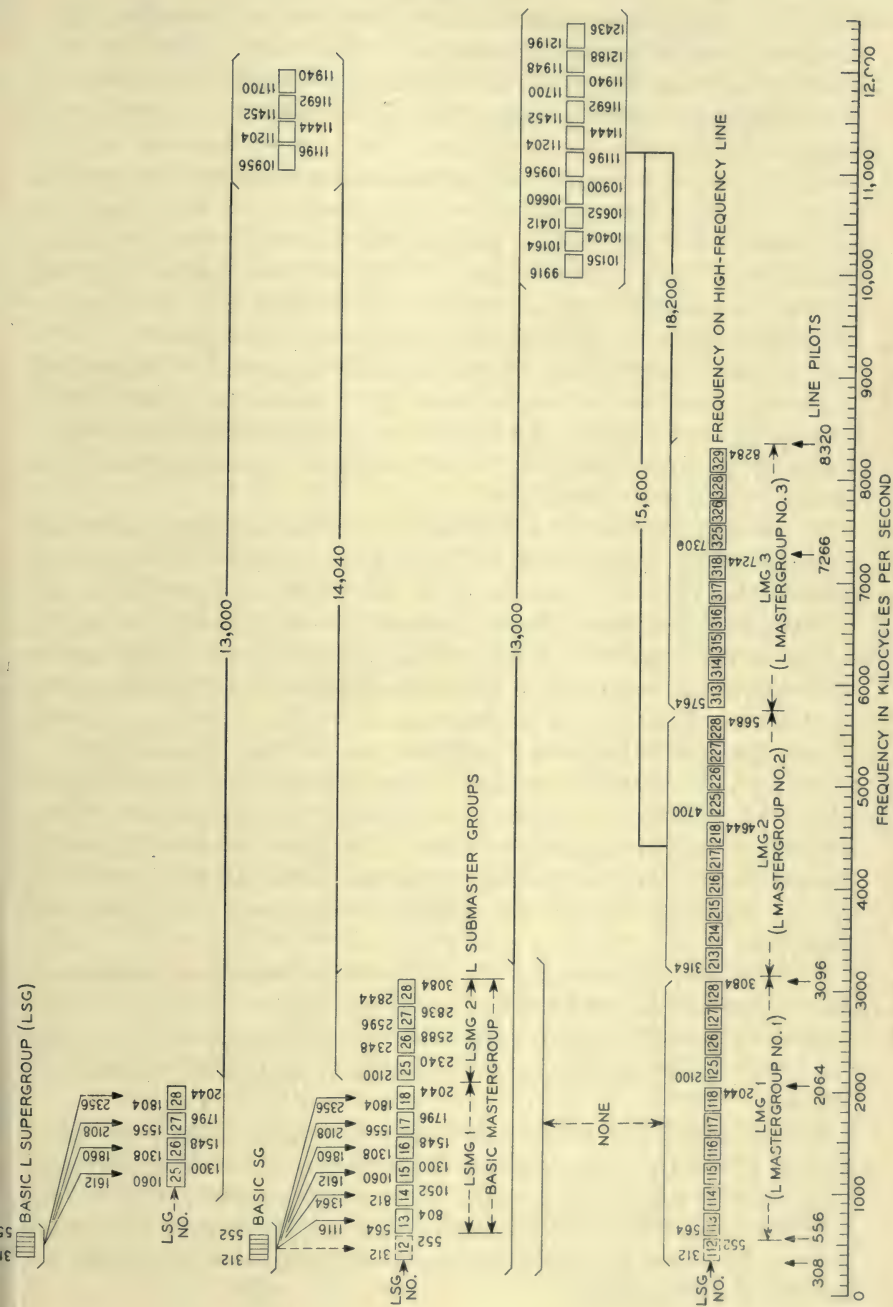
Mastergroup No. 1 is subdivided into two submastergroups. The lower six supergroups, comprising submastergroup one are modulated directly up from the basic supergroup located in the 312 to 552-kc band. The modulation and carrier supply equipment for these supergroups are the same units that are employed in the L1 system. The upper four supergroups comprising submastergroup two are obtained by modulating four supergroups located in the same frequency range as the top four supergroups in submastergroup one into the top part of mastergroup one.

The supergroup numbering system used for L3 has been adopted for easy identification of supergroups in their high-frequency positions. Each supergroup is given a three digit number. The first digit identifies the mastergroup, the second digit identifies the submastergroup, and the third digit identifies the L1 supergroup from which it was originally derived.

In the 1860 channel all-telephone allocation, supergroup No. 112, which corresponds to the basic L1 supergroup No. 2, may be used for high quality, long haul message circuits. When the system is used for telephone and television, supergroup No. 112 is restricted to circuits under 200 miles in length because of intolerable second order cross modulation between these signals and the television signal.

With these groupings of channels new modulating and carrier supply equipment is required for submastergroup No. 2 and mastergroups Nos.





2 and 3. The frequency allocation and modulation steps shown on Fig. 20 were chosen so that good performance could be obtained with relatively inexpensive filters used for suppressing unwanted sidebands and to develop lower sideband signals appearing right side up for transmission over the high-frequency line. The carrier frequencies used in the modulation steps were chosen so that some of the same filters could be used in the submastergroup two and mastergroup two and three modulators. The clear bands between the submastergroups and the mastergroups were chosen to permit the use of economical blocking filters at repeaters where circuits are to be dropped from the high-frequency line and to provide frequency space for the line pilots.

It will be noticed in Fig. 19 that facilities are provided for patching spare equipment into service for maintenance reasons or in the event of a failure in working units. Alarm features are incorporated in the submaster group and mastergroup units to indicate trouble conditions and initiate maintenance procedures.

The carriers required for the channel, group and supergroup units are supplied by equipment developed for the L1 system. The arrangements for supplying the new carrier and pilot frequencies are shown in Fig. 21.

A problem of primary importance in carrier supply design is the accuracy of the frequencies. There is both an absolute accuracy and a relative accuracy requirement. For transmission of some types of signals there is a requirement that the difference in frequency between a carrier at two terminals be less than 2 cycles per second. This extreme accuracy is achieved by using the oscillator at one terminal to control the frequency of oscillators at other points. A line pilot generated at the terminal in which the master oscillator is located is used as a reference frequency at points along the line where other terminals are located and by this means carriers are held to a relative accuracy of ± 1 part per 30 million.

In order that requirements for high quality television transmission may be realized on a 4,000-mile circuit, the output of the pilot supply must be extremely constant with both time and temperature changes. Deviations in the magnitudes of line pilots are maintained to less than 0.05 db.

Since a failure in the L3 carrier and pilot supply could cause interruption to service on an extremely large number of circuits, many precautions have been taken to make the equipment reliable. In addition, standby units, which are automatically switched in place of the regular units in the event of failure, are provided to improve the over-all reliability of the system.

The terminal equipment is normally mounted on standard 11' 6" duct-type bays. The submastergroup and mastergroup equipment required to handle 1,860 channels occupies two complete bays and portions of two others. One carrier and pilot supply is mounted in three bays.

3.6 POWER GENERATION AND TRANSMISSION

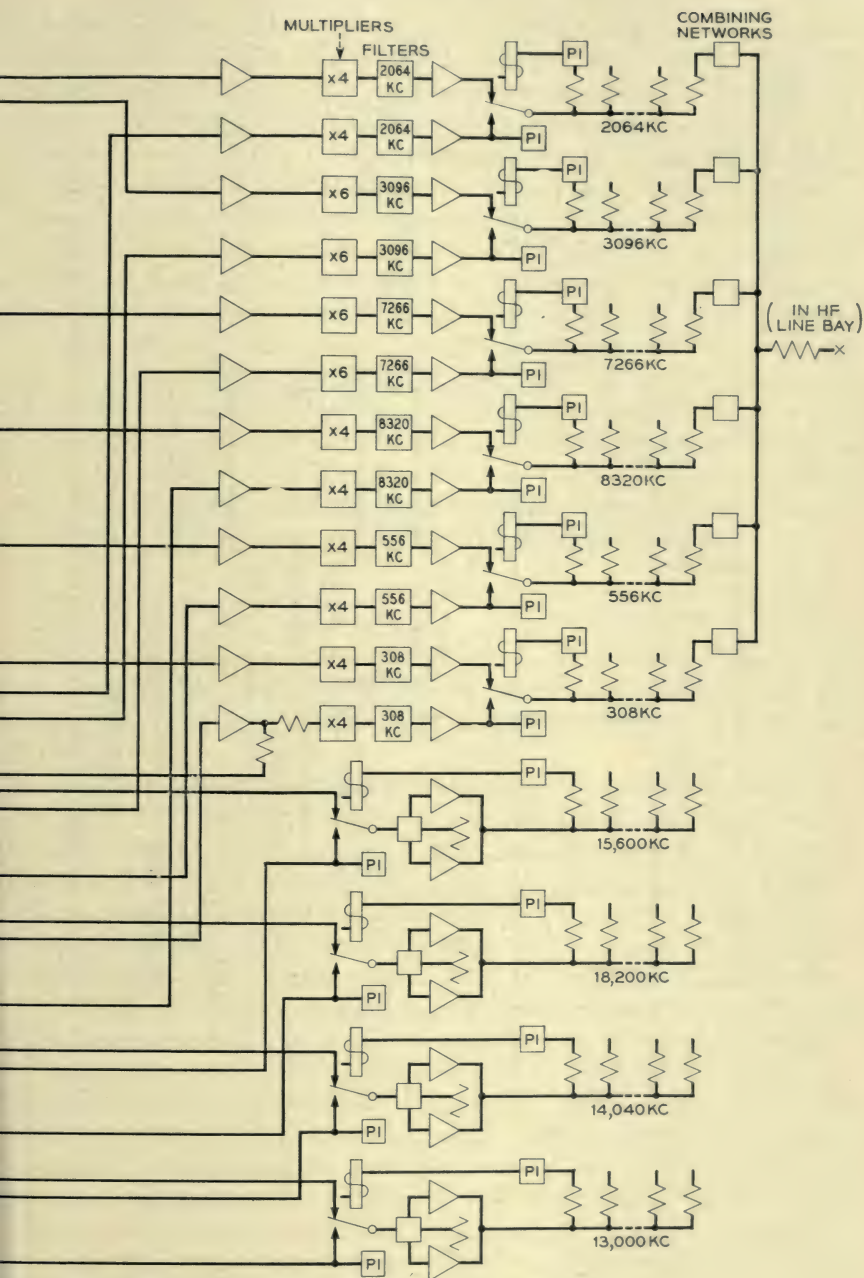
Power is transmitted to the auxiliary repeater over the inner conductors of each pair of coaxials as noted earlier. A maximum of twenty-one auxiliary repeaters can be fed from a main repeater. This limit is determined by the maximum potential the cable can safely withstand. Shorter spacings are dictated by geographical and plant layout considerations.

The power supplied to the coaxials at the main stations is generated by a motor-alternator set which consists of the alternator, an induction motor, a dc motor and its exciter, all coupled together on the same shaft. Normally, commercial power is used to drive the induction motor. When this source fails or the voltage goes out of prescribed limits the drive is transferred to the dc motor which operates from a 130-volt battery. If the commercial power is unusable for more than $2\frac{1}{2}$ minutes an emergency engine alternator is started and after a five minute warm-up period it replaces commercial power in driving the regular induction motor.

The constant current to the coaxials is supplied through a power control circuit which accurately regulates it to within ± 1 per cent of the desired value. A simplified schematic of the power control unit is shown on Fig. 22. The unit consists of two motor driven continuously variable transformers which supply power to the line transformer. The course control variable transformer is relay controlled and maintains the line current within ± 3 per cent of the prescribed value. The range of this transformer is sufficient to permit reducing the voltage to zero in order to turn down power on the system for maintenance purposes. The fine control variable transformer is regulated by an electronic regulator to maintain the line current within ± 1 per cent of the desired value.

The change in the line current in response to commercial power transients and transients introduced by changes in the motors driving the alternator requires careful consideration. By increasing the inertia of the motor alternator set with a fly wheel and carefully designing the frequency response of the above described power control regulator a satisfactory transient response has been obtained.

For the maximum length power section the potential applied to



ier and pilot frequencies.

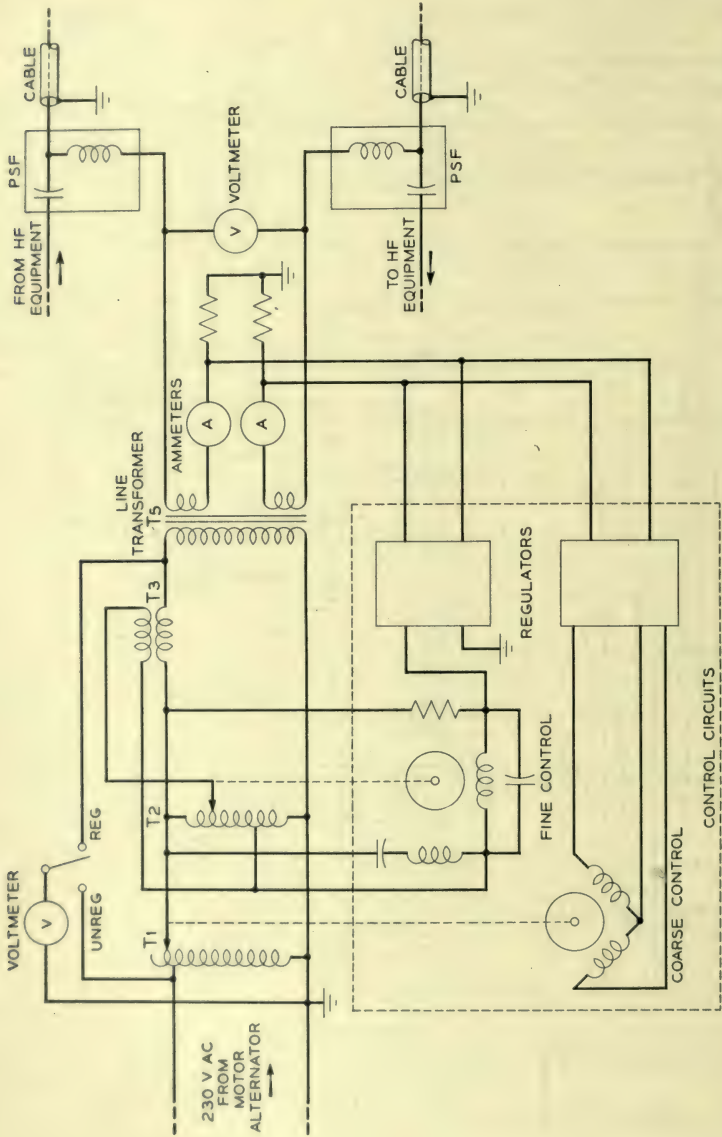


Fig. 22 — Simplified schematic of the power control circuit through which regulated power is fed to the coaxial circuits.

the cable at the main stations is about 2,000 volts rms from center conductor to ground. This potential diminishes about 100 volts per repeater section in going out from the power feed point or as the power section is shortened. (The maximum potential applied to the cable in L1 systems is about 800 volts rms between center conductor and ground). Extensive tests on the installed cable showed that corona develops in the cable at potentials varying in a random fashion between 1,200 and 1,600 volts rms. This would allow power feed points to be placed at a maximum spacing of about 100 miles. By replacing the nitrogen, with which the cables are normally filled, with a large molecule gas, sulfur-hexafluoride (SF_6), the corona potential of the cable is increased well above the maximum operating potential. Only the cable sections exposed to potentials greater than 1,200 volts will be filled with the new gas. Elaborate and thorough tests have demonstrated that no deterioration of the cables will result from the use of this gas. Some additional precautions are required in entering manholes and using high temperature torches for soldering when the SF_6 gas might be present.

3.7 ALARM EQUIPMENT

Since the auxiliary repeaters and certain of the main repeaters are unattended it is necessary that arrangements be provided to indicate at the attended stations when some piece of equipment fails to perform satisfactorily.

Auxiliary repeaters using pilot regulators are equipped with microammeter relays which monitor the operation of the regulator continuously. These relays provide an indication of the operation of the regulator and the power of the 7,266-kc pilot at the output of the repeater. When conditions change from the nominal by a specified amount the relay contacts close and are locked magnetically. This bridges an alarm pair in the cable and operates an alarm at the nearest attended repeater. By means of Wheatstone bridge measurements from this repeater over the same alarm pair, the repeater in trouble can be located and a maintenance crew dispatched to make the necessary equipment replacements. The relays can also be reset over the same alarm pair to aid in the location process or to clear alarms which were initiated at unaffected repeaters by deviations in the pilot due to troubles at preceding repeaters.

At main repeaters, microammeter relays are provided on all six pilots used to control the equalization of the system. Deviations in these pilots operate the automatic switching equipment and initiate the usual office alarms. Alarms are also provided to indicate fuse operation, transfers

from regular to standby equipment, and the condition of electron tubes in the terminal equipment amplifiers.

Provisions for connection to special alarm systems are made at main repeaters which are not fully attended. These systems extend the alarms to the nearest attended repeater and enable the attendant to determine in considerable detail the condition at the remote repeater. The attendant may also perform certain operations such as switching a working line to a spare line at the remote repeater.

3.8 MAINTENANCE

Maintenance of the L3 system requires equipment and methods for routine checking of the system and trouble location. Normally the auxiliary repeaters will be visited at intervals of about three months, when checks will be made of the power voltages and currents, the electron tube bias and change in bias with a fixed change in heater voltage (activity), and the pilot magnitudes. At these times amplifiers and regulators which fail to meet prescribed limits will be replaced, the 7,266-kc pilot will be brought to its normal value by adjusting the regulator gain, and the amplifier gain control in the output beta circuit will be adjusted by observing the 3,096-kc pilot.

For these routine tests and adjustments two portable test sets are provided. The power test set plugs into the repeater, amplifier, or regulator and provides for measuring power supply voltage and currents and electron tube cathode-grid voltages to an accuracy of ± 1 per cent. The pilot indicator makes it possible to measure the 7,266 and 3,096-kc pilots to an accuracy of ± 0.1 db.

For trouble locations at auxiliary repeaters a portable transmission measuring set has been designed. It is capable of measuring the power in a 500-cycle band at any place in the frequency spectrum from 50 to 11,000 kc to an accuracy of ± 0.5 to ± 0.02 db depending on its specific use and the care used in calibration.

At main repeaters, line sections will be checked for noise and modulation performance and equalizers will be adjusted at intervals of one week to several months. In addition, loss and gain measurements on sections of the office suspected of being in trouble will be made. For all general tests except equalization line up, point by point measuring equipment is provided, consisting of a 50 to 10,000-kc oscillator, the tuned transmission measuring set referred to above, a milliwatt power meter accurate to ± 0.035 db and a complement of attenuators, pads and comparison switches.



Fig. 23 — An engineer testing pilot transmission in an L3 repeater hut.

The adjustment of the manual gain equalizers to an accuracy of ± 0.02 db in a rapid and direct way is accomplished by special equipment described in the companion equalization paper.¹⁰ A visual gain and delay transmission measuring test set, capable of measuring gain to ± 0.05 db and delay to ± 0.02 microseconds, has been developed for observing the line performance and adjusting delay equalizers when the system is used for television.

A maintenance center is provided at about 200-mile intervals along the line to service the equipment removed from repeaters. At these points facilities are provided for the following: (1) electron tube testing; (2) regulator repair and adjustment; (3) transmission measurements on passive components; and (4) amplifier testing of sufficient scope to permit changing tubes and to determine whether an amplifier is suitable for further service in the line.

ACKNOWLEDGEMENTS

A system as complex as the L3 system is the result of a large scale cooperative development effort involving many Departments in Bell Laboratories. More than a hundred engineers and technicians have contributed to the design over a period of seven years. In the system planning aspects of the development covered by this paper particular mention should be made of the large contributions of L. G. Abraham, C. H. Bidwell and S. E. Miller.

BIBLIOGRAPHY

1. Abraham, L. G., Progress in Coaxial Telephone and Television Systems, Transactions of the A.I.E.E., **67**, pp. 1520-1527, 1948.
2. Crane, R. E., Dixon, J. T. and Huber, G. H., Frequency Division Techniques for a Coaxial Cable Network, Transactions of the A.I.E.E., **66**, pp. 1451-1459, 1947.
3. Holbrook, B. D., and Dixon, J. T., Load Rating Theory for Multichannel Amplifiers, Bell System Tech. J., **18**, pp. 624-644, Oct., 1939.
4. Bennett, W. R., Cross-Modulation Requirements on Multichannel Amplifiers Below Overload, Bell System Tech. J., **19**, pp. 587-610, Oct., 1940.
5. Mertz, Pierre, Data on Random Noise Requirements for Theater Television, S.M.P.T.E., J., **57**, pp. 89-107, Aug., 1951.
6. Baldwin, M. W. Jr., The Subjective Sharpness of Simulated Television Images, Bell System Tech. J., **19**, pp. 563-586, Oct., 1940.
7. Baldwin, M. W. Jr., Single Frequency Video Interference Data, Proceedings of the J.T.A.C., Annex 8, **2**, Dec., 1948.
8. Morris, L. H., Lovell, G. H., and Dickinson, F. R., The L3 Coaxial System — Amplifiers, see pp. 879-914 of this issue.
9. Rieke, J. W., and Graham, R. S., The L3 Coaxial System — Television Terminals, see pp. 915-942 of this issue.
10. Ketchledge, R. W., and Finch, T. R. L3 Coaxial System — Equalization and Regulation, see pp. 833-878 of this issue.
11. Mertz, Pierre, Fowler, A. D., and Christopher, H. N., Quality Rating of Television Images, I.R.E., Proc., **38**, pp. 1269-1283, Nov., 1950.
12. Dodge, H. F., Kruger, M. K., and Kinsburg, B. J., The L3 Coaxial System — Quality Control Requirements, see pp. 943-968 of this issue.

The L3 Coaxial System

Equalization and Regulation

By R. W. KETCHLEDGE and T. R. FINCH

(Manuscript received April 17, 1953)

The equalization and regulation problems of the L3 system are described and a theory of equalization of complex systems is outlined. The location and function of the various equalizers are explained including the roles and design of the various fixed, dynamic and manual equalizer networks. The analog computer used in the regulation system is described together with the cosine-equalizer adjusting technique used with manual equalizers. Finally the circuits and operation of the regulation system and its components are presented.

INTRODUCTION

Equalization is the process of correcting system gain and delay deviations sufficiently to permit the satisfactory transmission of signals. Regulation refers to that part of equalization which, by automatic means, corrects for relatively rapid changes in transmission. In the L3 coaxial system the transmission of television signals through more than 1,000 amplifiers introduces relatively severe equalization and regulation problems. Compared to the L1 system, the L3 system has nearly three times the bandwidth and over three times more stringent transmission objectives. Thus it has been necessary to devote considerable effort towards finding equalization methods that yield practical and economical solutions.

In previous systems it has often been the practice to design the bulk of the equalization system after the completion of an initial installation and the determination of the deviation characteristics of the system. In order to expedite the introduction of the L3 system into the field, equalization study and planning were initiated at the very beginning of the development of the system. One of the design methods was to make highly detailed studies using relatively inadequate data in order to find the major problems. As the data improved the studies were likewise

improved. While on the surface it might appear more practical to defer this work until the data are more adequate it has been found that these speculative studies are the only sure guide to ever getting the right data. Faced with such a complicated problem it is difficult to select from the vast amounts of things that might be important those relatively few items on which success or failure depends. In the L3 system there have been a number of critical problems which required intensive effort to find acceptable solutions, for example, the design of stable regulators to permit operation of 700 regulators in tandem, the choice of shapes for manual and dynamic equalizers and the selection of equalizer adjustment methods.

The design of equalization for the L3 system is not yet complete since the dynamic equalizer shapes are still subject to considerable uncertainty and the details of some of the final television mop-up equalization are yet to be settled. However no major difficulties are anticipated in equalizing the telephone system to be installed from Philadelphia to Chicago during 1953. Although amplifiers having the final gain characteristic were operated in the trial line between New York and Philadelphia for the first time in November, 1952, it was immediately possible to transmit quite satisfactory television pictures over the 200-mile loop using existing equalizers. It therefore appears that equalization will not limit the rate of field installation of the L3 system.

THE PROBLEM

The 4,000-mile coaxial cable has a gain distortion of nearly 40,000 db between 0.3 and 8.5 mc. Although the amplifiers reduce the distortion to perhaps 200 db, (and 100 microseconds), they leave a residue characteristic that is considerably more difficult to equalize. Further it is necessary to deliver service to intermediate offices spaced on the average about 120 miles apart. This requires equalization of high precision at numerous intermediate points. A further problem is the variability of the transmission characteristic due to manufacturing deviations plus time and temperature changes. Also, gain distortion cannot be permitted to exceed about 5 db at any point in the line or the signal misalignment will result in degraded signal-to-noise ratios.¹

The overall transmission objectives¹ are of the order of 0.25 db and 0.1 microsecond which, if allocated among the over 1,000 amplifiers, lead to rather unrealistic amplifier requirements. In fact, individual amplifiers do not always meet the 4,000-mile overall requirements. Thus it is the problem of the equalization designer to provide a mop-up sys-

tem that will permit attainment of the transmission objectives at all service points and at all times.

EQUALIZATION THEORY

One of the steps in the solution of the general problem has been to develop a "theory" of equalization. This theory merely applies information concepts to the equalization problem to determine what information is required, when it is needed and how it may best be used. This theory has stimulated the development of novel equalizer adjustment techniques and has been of assistance as a guide to the attack on the general problem.

In order to equalize a system the man or machine who is to perform the action must know what corrective steps are required, and for this he must have some kind of information as to the present state of the system and as to the desired state. Second, he must have the necessary tools to convert the system from its present state to the desired state. Consider the first problem, the determination of the corrective steps required. The problem is to determine what we need to know, when we need to know it and especially in what form we are able to utilize the information most efficiently.

We can assume we know the desired state of the system; which is usually a constant loss with constant delay over the frequency range of interest. As to the present state of the system we note that sufficient information can never be obtained to equalize a system perfectly because of the finite bandwidth of the system and because the system changes with time. This is not a new fact, nor apparently a very important fact, because the system need not be perfectly equalized for satisfactory transmission of signals. It leads, however, to the converse idea, which is important — namely, that out of this infinite amount of information regarding the state of the system one should collect only the minimum amount that is needed. This implies making no more measurements of the state of the system than are absolutely necessary to perform the correction to a degree permitting satisfactory transmittal of the signals. The main purpose of this is, of course, to economize on time and effort required to obtain information, but it should be noted that excess information may be a source of confusion to the equalization operator.

To put this in the form of a rule, we have:

Rule I

Collect only that minimum of information as to the state of the system as will permit equalization to the required degree for satisfactory transmission of the signals.

Having established that a minimum of information should be collected, we return to the time variation of the system, which in theory can make the information obsolete before it can be used. However, without yet bringing in the practical fact that its most rapid rate is relatively slow, we should introduce the fact that the ways in which the system can vary at its most rapid rate are quite restricted as compared with the manner in which it can vary at slower rates, and we note that even these are relatively limited. (This probably applies to most transmission systems, not just to the L3 coaxial.)

Thus while the information regarding the more rapid, but simple changes must be collected more frequently it consists of a small amount of information per sample; whereas the information regarding the slow (but more complex) changes need not be collected very often, but it represents a relatively large amount of information per sample. Since the total rate of information collection is proportional to the information per sample times the rate of sampling, the minimum information collection principle would say that the rate of sampling should also be held to a minimum.

This demonstrates the value of association of particular types of system change with the rate and amount of their variation, because one may thus eliminate from the more rapid sampling the collection of information about changes that occur at slow rates. Furthermore, one may establish the sampling rates for the various system effects at the lowest possible value. (There is also a very practical value in knowing the rates and amounts of the various deviations, because system misalignment requirements force the equalization to be suitably distributed along the line.)

In the form of a rule, this is:

Rule II

To the greatest practicable extent the overall system behavior should be separated into individual effects each having its own time rate of occurrence and corrections should be made for each effect at the minimum tolerable rate for each.

It is of interest to note that, if we couple the logic of Rule II with the fact that the fastest changes (due to changes in the temperature of the repeater huts) take hours to become appreciable, we see that the continuous collection of information from continuous pilots (and the continuous correction by pilot-controlled regulators) is in principle unnecessary and inefficient — except, of course, for its other function of giving alarms under trouble conditions.

The technical problem of determining the equalization states of the system is normally solved by sending some kind of signals over the system and observing the effect of the system on those signals. The raw data are usually in the form of loss and delay as a function of frequency. On the basis of these data, the equalization operator desires to correct the system by means of some equalizers which have adjustable transmissions and delays as a function of frequency. Thus the operator has a group of controls to be operated plus some data which has encoded in it the information as to the proper adjustment of each control. From these data, and a knowledge of the effect of each control, the operator must suitably compute the proper adjustments. As this may be too complicated a process to attempt on a trial and error basis, (or by numerical methods), it is quite an obvious advantage to the operator to receive the data as to the state of the system, not in its original form, but in the form of the necessary adjustments to his equalizer controls. This new form of the data simply represents a decoding process based on the available controls. An operator with the same data, but different and perhaps more complicated equalizers, would need the data in a form suited to his different equalizers.

Consequently:

Rule IIIa

The information as to the state of the system may best be presented to the equalization operator in the form of the necessary adjustments of the available equalization controls.

This rule has a closely related corollary which is based on the fact that the available equalization controls determine the amount of information that is needed. For example, if the independent gain equalization controls are " n " in number, measurement of the gain of the system at " n " suitably chosen frequencies is sufficient to determine the settings. (If the controls are not independent, fewer than " n " frequencies need be measured.) This is, of course, a restatement of the fact that " n " unknowns may be determined by solution of " n " independent simultaneous equations. The unknowns are the equalizer settings and the simultaneous equations are the relationships of the shapes controlled by each equalizer to the total system error.

Thus:

Rule IIIb

In general, the necessary and sufficient condition for the determination of " n " independent equalization control settings is the knowledge of the system's

equalization error at "n" independent frequencies plus the knowledge of the effect of each of the "n" controls at each of the "n" frequencies.

It should be noted that this statement of the rule assumes analysis by frequency rather than by transient behavior. This approach is used because, at present, equalizers are usually designed on a frequency characteristic basis. The validity of the rule is however more general.

From these two rules we can derive another regarding the minimum information principle that is similar to Rule I but is actually quite independent of it.

Rule IV

No more information should be gathered from the system than is necessary to provide sufficiently accurate control setting information for the equalization operator.

In this case the superfluous information may actually cause confusion or harm. It will, at the very least, confuse a manual operator to know of an error he is powerless to correct, whereas a mechanized system, on the other hand, would probably go berserk if it obtained too much information.

It is evident that the minimum amount of information that would be obtained in accordance with Rule IV would be the same as that obtained in accordance with Rule I only if the design of the equalization were optimum, because then the equalizer shapes (and the number of shapes) would just suffice to permit satisfactory transmission of the signals.

Up to this point we have determined in general what information is needed, when it is needed and the optimum form of its presentation. Now let us proceed to examine what to do with the data; which involves the nature of the equalization operator as well as his equalization tools.

If we followed Rule II rigorously, we should have several different type of controls; one for each of the effects having different time rates of occurrence. For purposes of illustration, however, we need assume only two rates — one quite rapid and the other very slow. It will be postulated here that very rapid equalization operations are most economically performed by machine, such as for example, the automatic regulation for cable temperature variations. Likewise relatively infrequent adjustment will be assumed to be best performed by a suitably informed human operator.

The first principle to note is that the only real distinction here is the rate at which data should be refreshed and acted upon. In either case

the data should be in the same form; a set of numbers (or their equivalent) representing equalizer changes.

Probably the most useful result of this theory of equalization has been the conclusion that mechanization of the equalization process, particularly in regards to computational techniques, permits substitution of simple logical methods for inefficient trial-and-error adjustment processes. This led to the use of an analog computer in the regulation system and, for manual equalizers, the development of measuring circuits that read directly in terms of equalizer adjustment error.

EQUALIZATION

LOCATION AND FUNCTION OF EQUALIZERS

The location of equalizers in the L3 system when only telephone is transmitted is shown in Fig. 1(a). Combined telephone-television equalization is shown on Fig. 1(b). When telephone and combined systems use the same spare line, that line is equipped for television. The general features of the main-repeater layout have been described in a companion paper.¹ The detailed layout of equalization is designed to meet the requirements of both telephone and television service and the need for flexibility in television network arrangements. In addition, switching of telephone or television service to a spare line must not appreciably degrade service.

Switching sections longer than 120 miles must be provided with an intermediate step of equalization to prevent excessive signal misalign-

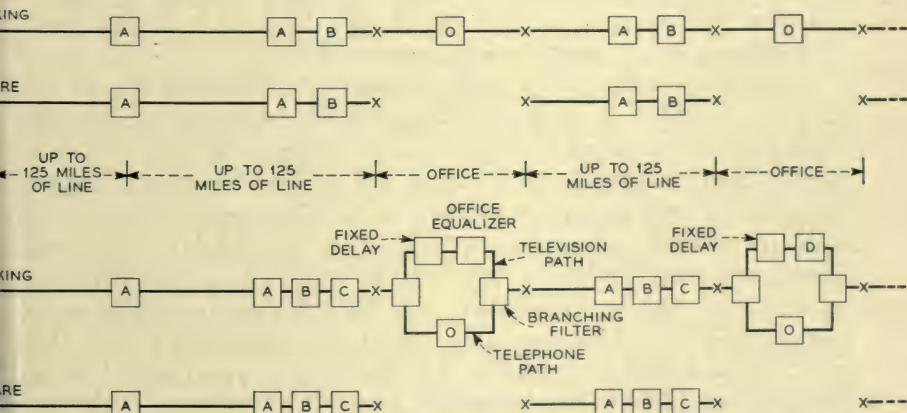


Fig. 1 — Typical equalizer locations in (a) telephone and (b) combined systems.

ment. This intermediate point is referred to as an equalizing auxiliary repeater. It is provided with the so-called A equalizer consisting in turn of fixed, manual and automatic gain equalizers. It reduces the gain error to less than 0.5 db using the fixed and manual sections and prevents appreciable degradation of this residue during an ensuing three month period by the action of three "office" regulators using pilots at 308, 2,064, and 7,266 kc controlling three regulating networks. One of these networks is the $\sqrt{f}$ shape of the receiving amplifier.² The other two are first order corrections for vacuum tube aging and repeater temperature changes.

At offices, (switching main repeaters), on telephone systems further equalization is required to permit line switching of telephone channels and special service signals such as telegraph. Also it is not practical to provide telephone equalization on a cumulative basis over longer links than a switching link because of frogging and dropping.¹ No telephone channel rides for more than 800 miles in the same frequency location and dropping breaks up the pattern still further. Thus, for telephone, the switching links, which may number between 30 and 40 in a long system, are independently equalized.

The B equalizer contains manual and automatic sections, the latter being three regulating networks omitted in the A equalizer. Thus an A plus a B forms the final telephone equalization. The residue of five, independently-adjusted AB links must meet telephone requirements without frogging and 30 to 40 links must meet these requirements with 800-mile frogging. This performance must continue to be met in the presence of a normal amount of spare line switching. Also it is very important that the character of the residues be such as not to throw an undue burden on the television equalization.

In combined systems the more stringent requirements require the addition of further manual equalization to the switching section. The residue at the output of a C equalized switching section must be sufficiently small that a 4,000-mile circuit will continue to meet television transmission requirements in spite of a normal amount of spare line switching. Also the C equalizer in conjunction with the AB must permit several switching links to be connected in tandem without further line equalization. The C equalization also contains an adjustable delay section which in conjunction with a fixed delay equalizer in the office path provides delay equalization for the television part of the band, 3.6 to 8.5 mc. This adjustable section builds out the line to match the fixed unit.

The A-B-C pattern is for equalization of individual switching lines

and these equalizers are adjusted on a single switching link basis. On the office side of the switches are system components that also require equalization. In the telephone system these include such things as office cabling and hybrid coils. Also there is an office flat loss of nearly 30 db. Thus the lines are, in effect, operated to give a 30 db gain while the offices give a 30 db loss. Aside from the flat losses there are distortion shapes but fortunately these are all well approximated by a $\sqrt{f}$ shape and a single manually adjustable $\sqrt{f}$ equalizer plus suitable choice of flat loss provides adequate telephone office equalization. This is referred to as the O equalizer. It is adjusted to make the particular office have a flat characteristic.

In the office circuits of combined systems are branching filters to separate the telephone and television bands. The O equalizer is reused in the telephone path. The television path includes the fixed delay equalizer and a manual gain equalizer to correct for office cables, hybrids etc. After the tandem combination of several independently equalized lines and offices further equalization will be required. This will be accomplished by a multi-control manual D equalizer, having both gain and delay sections, inserted in the television only path at approximately 400-mile intervals. These D equalizers will be used to form 800-mile pilot links which are independently equalized to a degree permitting putting any five such links in tandem to form a 4,000-mile circuit without further equalization.

FIXED EQUALIZERS

The line amplifier can properly be considered as the first step of fixed equalization.² Also acting at this same level are artificial cable networks used to build out the repeater spacing to 4 ± 0.2 miles, as well as the basic equalizer of the amplifier to take up differences between cable types. These devices are described in a companion paper.¹ The final step of fixed equalization is the so-called "design deviation equalizer" associated with all A equalizers. This equalizer comes in two versions, one for use with sections containing 23 to 32 repeaters and one for use in sections of 10 to 22 repeaters. In those few cases of less than 10 repeaters the fixed equalizer is omitted.

The function of the design deviation equalizer is, first, to correct for the design error of the average repeater and second, to recenter the manual (cosine) equalizers. Although the average repeater matches its four miles of cable to within ± 0.12 db this design error accumulates to over 3 db in 30 repeaters and further the shape is a difficult one to equalize. In order to keep the number of designs to a minimum only two sizes

are used, 19 and 28 repeaters, and the residue is corrected by the manual equalizers. This residue may be positive or negative depending on whether the fixed equalizer over or under compensates. Thus there is only a small tendency for these residues to accumulate in long systems. However the inaccuracies of match between the fixed equalizer and the gain of the average repeater section tend to accumulate systematically and must therefore be kept small. This brings out the importance of statistical quality control of amplifier manufacture² since any systematic shift in the amplifier gain characteristic will accumulate and may consume excessive manual equalizer range or may lead to excessive equalization errors. For example, a shift of only 0.05 db in the gain of the average amplifier would represent 1.5 db in 30 repeaters, 50.0 db in 1000 amplifiers and thereby becomes an extremely serious matter. Thus quality control of the amplifier is a vital part of the solving of the equalization problem.

The various effects that consume the range of the manual equalizers produce for some shapes an unsymmetrical consumption of range. If uncorrected this would produce larger range requirements in the manual equalizers as well as introduce new shapes to be equalized. The manual equalizer shapes are symmetrical and their errors cancel if equal amounts of positive and negative range occur in the system. Any systematic offset of a particular shape tends to introduce new shapes due to the manual equalizer networks themselves. By appropriate modification of the shape of the fixed equalizer it is possible to recenter the manual equalizers so that on the average the manual shapes are in the center of their range.

DYNAMIC EQUALIZERS

Any long transmission system suffers from relatively rapid gain changes and in the L3 coaxial system, as in many previous systems, the necessary corrections are performed automatically by pilot controlled regulators. Pilot tones are transmitted over the line at a reference level and, at appropriate points, regulators pick the pilots off the line, observe the deviation in pilot levels from the reference values and restore the pilots to or very nearly to the reference values by the use of regulating networks.

There are fundamentally two causes of fast gain changes, time and temperature. Time produces vacuum tube aging and in spite of their feedback the line amplifiers change gain. In one week a 4,000-mile system is expected to change by as much as 5 db due to the aging of the 6000 tubes or so in the transmission path. Because of different thermal

characteristics, temperature affects cable and repeaters semi-independently. In a 4,000-mile system one week is expected to engender as much as 8 db gain change by change in repeater temperature. Cable changes can exceed 100 db per week. Thus these effects must be corrected to a very high order of precision to maintain good television or telephone service.

These gain changes are not the entire story; when the gain changes in the band it also changes outside the band and usually by an even larger amount. Thus equalization of the in-band gain leaves outband (above 8.5 mc) gain changes which produce in-band delay changes of several microseconds. Because of the difficulty and complication of providing automatic delay equalization it is necessary to equalize these delay changes on a gain basis, by at least partial correction for outband gain changes. Further, since satisfactory pilot transmission is possible only within the band, these out-band changes must be predicted from the in-band gain changes. This effect, in itself, indicates the use of equalizers whose individual shapes are those produced by specific system causes producing a correlated change in many elements. Thus building the regulating networks to match the effects of the individual system causes and matching to 10 or 12 mc rather than just to 8.5 mc permits simultaneous gain and delay equalization. Such a set of shapes is also more accurate because it is matched to the special ways in which the specific system can change rapidly.

In theory the regulating networks could be built to match linear combinations of the cause shapes but there are two difficulties. First, not all of the cause shapes are known accurately or often even roughly at the introduction of a new system into the field. The mixtures of shapes cannot be determined without the missing ingredients. Second, a system is not a static design. Experience suggests improvements and thus occasions will arise where one will want to change the correction for a particular cause. If the networks represent mixtures of the causes this necessitates changing all the networks. If specific networks match specific causes only the appropriate network needs replacement.

The Computer

The use of cause shapes leads to a problem to which the computer provides the solution. These cause shapes are broad effects covering the entire band and more. Thus no one pilot is a measure of a specific cause. However by a process equivalent to the solution of simultaneous equations the pilots determine the amounts of an equal number of cause shapes that will restore the pilots to normal. Thus the computer trans-

lates pilot errors into shape errors and drives the appropriate regulating networks to obtain the corrections.

Let the equalizer shapes be given by functions of the form

$$S_n(f) = k_n F_n(f), \quad (1)$$

where " n " is the subscript number identifying the particular equalizer.

(The capital " N " is reserved for the total number of equalizers.)

" $F_n(f)$ " is the equalizer shape (on a "unit basis") as a function of the frequency " f ".

" k_n " is the amount of shape introduced by adjustment, " k_n " may be positive or negative.

" $S_n(f)$ " is the resultant shape put in the system by adjusting $F_n(f)$ by an amount k_n .

The total shape introduced by all " N " equalizers is obviously;

$$S_{\text{total}}(f) = \sum_{n=1}^{n=N} S_n(f) = \sum_{n=1}^{n=N} k_n F_n(f). \quad (2)$$

To obtain a match of S_{total} to the given equalization error, S_{given} , at " M " frequencies from $m = 1$ to $m = M$, requires that;

$$S_{\text{total}}(f_m) = S_{\text{given}}(f_m). \quad (3)$$

at each frequency from f_1 to f_M . Or, in terms of equation (2)

$$S_{\text{given}}(f_m) = \sum_{n=1}^{n=N} k_n F_n(f_m) \quad (4)$$

again, at each frequency from f_1 to f_M .

All of the important conclusions regarding the action of an equalization computer are implicit in the " M " equations indicated by equation (4).

Consider a case where there are three shapes. Let the information as to the difference between the system state and its desired state be determined by the deviation of three pilot levels which are observed to be δ_1 , δ_2 and δ_3 at the pilot frequencies f_1 , f_2 and f_3 . The problem is to find the values of k_1 , k_2 and k_3 that will give a match at these frequencies. This means that the following equations must be satisfied:

$$S_1(f_1) + S_2(f_1) + S_3(f_1) = \delta_1 \quad (5)$$

$$S_1(f_2) + S_2(f_2) + S_3(f_2) = \delta_2 \quad (6)$$

$$S_1(f_3) + S_2(f_3) + S_3(f_3) = \delta_3 \quad (7)$$

Thus

$$k_1 F_1(f_1) + k_2 F_2(f_1) + k_3 F_3(f_1) = \delta_1 \quad (8)$$

$$k_1 F_1(f_2) + k_2 F_2(f_2) + k_3 F_3(f_2) = \delta_2 \quad (9)$$

$$k_1 F_1(f_3) + k_2 F_2(f_3) + k_3 F_3(f_3) = \delta_3 \quad (10)$$

The solutions for k_1 , k_2 and k_3 take the form

$$k_n = a\delta_1 + b\delta_2 + c\delta_3, \quad (11)$$

where a , b and c depend solely on the shapes $F_m(f)$.

Thus the values of the δ 's may be decoded into the values of the equivalent k 's by the simple process of multiplying each " δ " by some fraction that is a function of the equalizer shapes; or, more precisely, by a fraction that is a function of the values of the various shapes at the frequencies f_1 , f_2 , etc.

The circuit of the computer is quite simple; consisting of about N^2 resistors for the control of " N " networks by the deviations that are measured at $M = N$ pilot frequencies. This simplicity is valuable for its own sake, but, as previously noted, there is considerable value in the fact that substitution of new equalizer shapes requires only that changes be made in the values of some of the resistors.

Fig. 2 illustrates the principle by showing the computer circuit required for the previous example of three shapes for which information is given at three frequencies, f_1 , f_2 and f_3 . The three dc voltages representing the deviations δ_1 , δ_2 and δ_3 are decoded by simply cross-connecting them to the three pairs of output terminals through fixed resistors chosen to satisfy the relationships of equations (8), (9) and (10). The output voltages will be proportional to the desired equalizer correction quantities k_1 , k_2 and k_3 .

In general the calculation of some of these resistors will give negative values. Thus it will generally be necessary to require that the dc voltages representing the errors δ_1 , δ_2 , etc., be available in both polarities. Alternately, the errors can be provided in only one polarity and the circuits to which the computer outputs connect can provide the push-pull circuit. This latter course has been used in the L3 regulators as indicated on Fig. 2.

The effect of using the computer is as follows. If one pilot changes, all regulating networks correct but in such proportions and polarities as to produce no gain change at any pilot frequency except at the one originally disturbed. If all pilots deviate in proportions corresponding

to one of the shapes, no network corrects except the one corresponding to the original pattern of pilot deviation.

If conventional regulator circuits were used with particular pilots assigned to particular shapes the interactions would be intolerable. Thus the computer removes the restrictions on the choice of shapes imposed by regulator interactions. In turn this permits freedom in changing shapes as system data improves. It is important to provide the initial equalizers before adequate data on the system are available. As the system grows in length the equalization must be improved but the data improves also. Thus there is an economic benefit in providing a flexible equalization plan which allows the earliest possible commercial use of the system.

Dynamic Equalizer Requirements

Because temperature and aging variations can result in both gain and loss variation with respect to the average transmission, the equalizers must be capable of inserting both loss and gain compensation. Also, it is extremely desirable that equal gain and loss settings for the equalizer result in symmetrical transmission characteristics with respect to the average. In a long transmission system, some of the sections will insert excess gain and others, excess loss. If the equalizer gain and loss characteristics are symmetrical, the residue will be related to the system

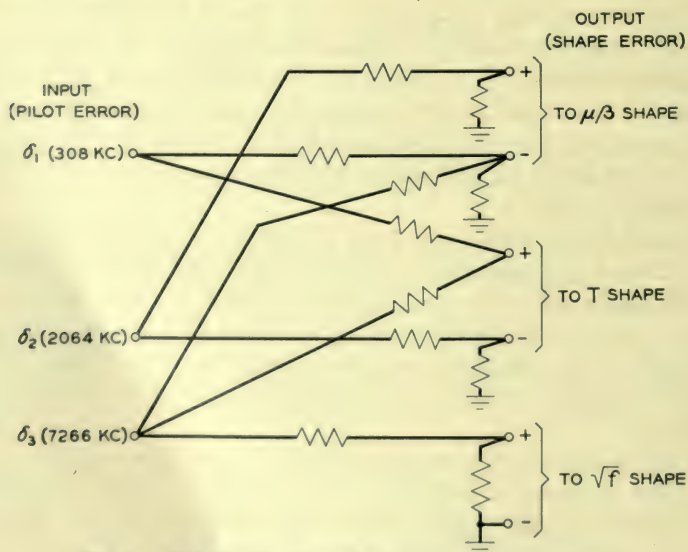


Fig. 2 — Schematic of regulation system computer.

deviation and some constant part of the equalizer characteristics. If this is not the case the residues can produce new variable deviations shapes and thereby lead to increased complexity in following stages of equalization.

As previously described, equalization of both gain and loss is required and this implies active equalizers. Although a number of methods were studied, noise and modulation requirements led the L3 system to the so-called "block of gain-block of loss" design. The equalizers are passive networks and in order to realize both gain and loss adjustability, the normal setting loss must at least equal the total amount of gain adjustment. The various equalizers are combined into two to four groups whose loss is compensated by corresponding numbers of flat gain amplifiers.²

The required number of such blocks of loss and gain depends upon the amount of system gain variation to be equalized. The determination of the shapes and magnitudes of system variations is an important system problem. Large amounts of study are necessary in order to evaluate the system sensitivity to various changes; the determination of the magnitude of these causes, such as temperature variation, aging rates, etc.; and the determination of maintenance intervals that provide an economic balance between maintenance expense and system cost. These must be studied in detail to provide the equalizer designer with shape and range data for his dynamic equalizer designs. The equalization characteristics and maximum ranges for the two most important dynamic equalizers aside from cable temperature, namely, repeater temperature (T) and vacuum tube aging $\mu\beta$ are shown on Fig. 3. In order to maintain satisfactory transmission during a maintenance interval, the dynamic equalizers must be able to match any characteristic within the maximum ranges and throughout the transmission band to within 1 to 2 per cent. As noted previously the necessity for simultaneous delay equalization requires the dynamic networks to also make at least an approximate correction in the out-band region. This is shown on Fig. 4.

The control element in the equalizers is a thermistor whose available resistance range is 30 to 1050 ohms. ($\sqrt{f}$ of line amplifier^{1, 2} 125 to 2000 ohms). Thus the regulation ranges shown in Fig. 3 are realized using this one variable resistance element in each equalizer.

The following is a summary of the dynamic equalizer requirements.

1. Provide symmetry in regulation characteristic.
2. Minimize flat loss.
3. Match prescribed gain variation to a high degree of accuracy within transmission band.

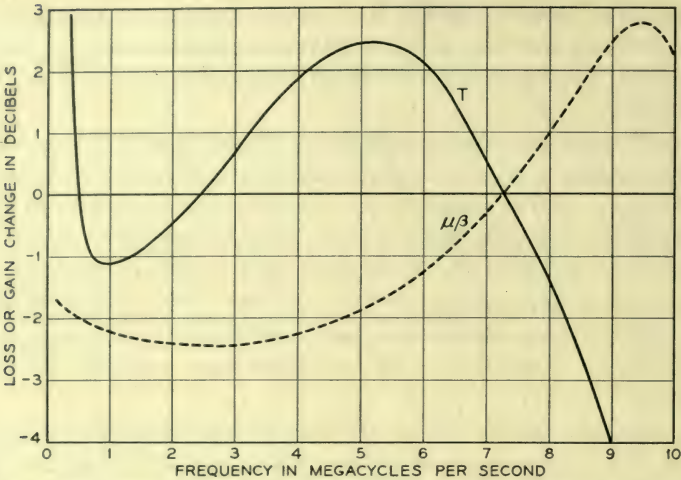


Fig. 3 — Regulation shapes and ranges for tube aging (μ/β) and repeater temperature (T). The curves given can occur as either gain or loss and apply to 30 repeaters.

- 4. Provide satisfactory equalization of in-band delay distortion due to out-band gain change.
- 5. Each equalizer controlled by a single variable resistance element.
- 6. The desired performance to be realized in 75-ohm circuits.

Dynamic Equalizer Design

The design used is the structure commonly known as the Bode Regulator,^{3, 4} and shown in block schematic form in Fig. 5. An ideal regulating

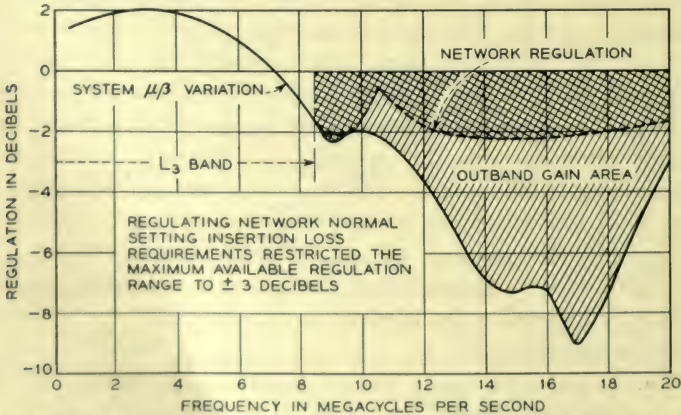


Fig. 4 — Match to out-band gain change to reduce in-band delay distortion.

network would insert a deviation characteristic

$$\theta = f(R_T) \cdot f(\omega) \text{ db.} \quad (12)$$

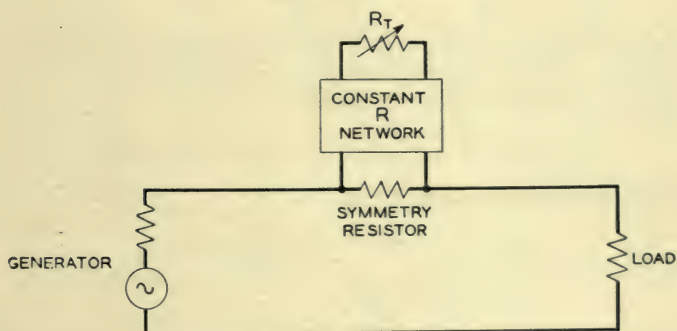
Such a network would yield a given shape independent of setting and thus would have linearity as well as symmetry. Actually the Bode network has symmetry but only approximates linearity. As shown, it may be designed to insert a variable impedance either in series or in shunt with the line. When using only one control element it is not a constant resistance structure and must be operated between terminations of the design value. The choice of series or shunt configuration was governed for L3 by the fact that very low thermistor resistance would require excessive dc currents from the regulators. For a 75-ohm circuit and this thermistor limitation the series regulator provides lower normal-setting loss.

The insertion gain for the series regulator may be written as:

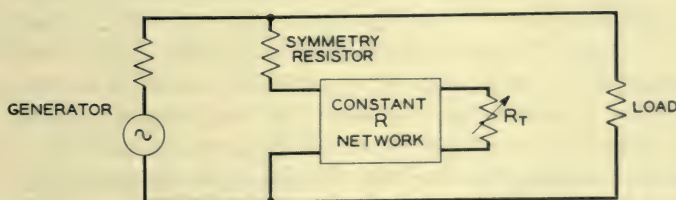
$$\theta = \theta_N + 2 \operatorname{arc} \tanh \left[\frac{R_1 - R}{R_1 + R} \cdot \rho e^{-2\Gamma} \right], \quad (13)$$

where θ_N = The normal setting flat loss as given by

$$\alpha_N = 20 \log \frac{R_1}{R} \text{ decibels.} \quad (14)$$



(a) SERIES TYPE



(b) SHUNT TYPE

Fig. 5 — Block diagram of Bode Regulating Network.

R_1 is the symmetry determining resistor and is in parallel with the input terminals of the shaping network in order to limit the maximum value of the series impedance. The equation of symmetry may be expressed as

$$R_1^2 = \left(1 + \frac{R_1}{2R_0}\right)R^2. \quad (15)$$

R is the image impedance of the shaping network. The thermistor has this value at its normal setting.

R_0 is the impedance of the circuit in which the regulator is inserted, in this case, 75 ohms.

ρ is the reflection factor between the network impedance, R , and the thermistor impedance, R_T and is given by

$$\rho = \frac{R_T - R}{R_T + R}. \quad (16)$$

Γ is the image transfer constant of the shaping network and may be expressed as

$$\Gamma = \sigma + i\psi. \quad (17)$$

Returning to expression (13) for the insertion gain, if the series expansion of the arc tangent is used and the higher order terms discarded, the insertion transfer constant becomes

$$\theta = \alpha_N + Ke^{-2\sigma} \cos 2\psi + iKe^{-2\sigma} \sin 2\psi, \quad (18)$$

or the insertion loss may be expressed

$$\alpha = \alpha_N + Ke^{-2\sigma} \cos 2\psi \text{ nepers}, \quad (19)$$

$$K = 2 \cdot \frac{R_1 - R}{R_1 + R} \cdot \rho. \quad (20)$$

This approximation is valid for design purposes if the regulation range is not too large or the accuracy required too great. For example, omission of the cubed term in the repeater temperature design produced a maximum error of 0.014 db and in the tube aging design, 0.007 db.

From the above expression it is seen that both real and imaginary parts of the image transfer constant for the shaping network enter into the insertion loss expression, and thus the relationship between the desired insertion loss and the shaping network makes the design process difficult. At this stage in the design considerable art and ingenuity is required in order to continue the design efficiently. The ap-

proach taken depends upon the regulation characteristic desired and the experience of the designer. Since both loss and phase are involved, familiarity with loss-phase relations is important. After a design has been blocked out, repeated modifications of the network parameters usually indicate the adequacy of the configuration chosen. Usually several sections in tandem are required in the shaping network in order to obtain precision of match, and some of these most likely will be all pass sections in order to control the phase ψ independently of the real part σ . It may be noted that at cross over points in the regulation characteristic, i.e., zero regulation, either the phase ψ has to be 45 degrees, or odd multiples thereof, or the loss σ has to be infinite. Usually the

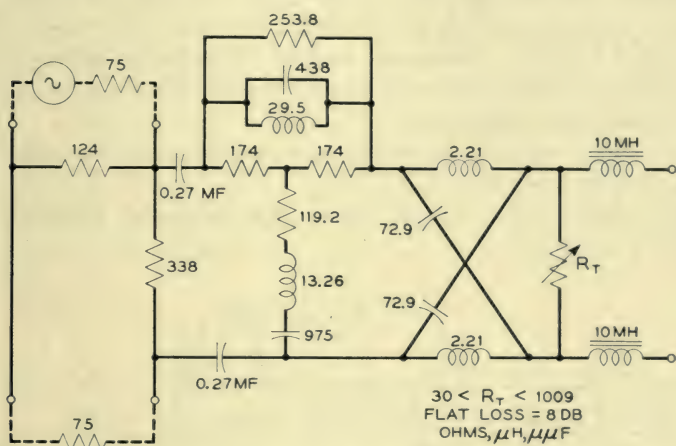


Fig. 6 — Circuit of the regulating network for repeater temperature.

phase is made the controlling term in both zero and peak regulation points.

The designs presently used on the L3 system for the repeater temperature and tube aging equalizer are shown in schematic form on Figs. 6 and 7. The 0.27 microfarad capacitors and 10-millihenry inductors are employed in order to supply the dc heating current to the thermistor. The shunt resistors placed across the 75-ohm line, in one case 124 ohms and in the other case 133.4 ohms are used in order to make the left side driving point impedance equal 75 ohms at normal setting of the thermistor. These networks are used in cascade with the manually adjusted, constant resistance networks and it seemed desirable that the impedance characteristic of the dynamic equalizers be relatively good at one pair of terminals.

MANUAL GAIN EQUALIZERS

One of the more difficult equalization design choices is the selection of shapes for the manual equalizers. In the L1 system some of the shapes used were so-called "bump" shapes. This type of shape reduces the amount of shape overlap and thereby tends to reduce the adjustment problem when conventional adjustment methods are used. For the L3 system many types of shapes and circuits were carefully studied. The computational concepts of equalizer adjustment resulted in the consideration of shapes that would otherwise be impractical.

The shapes finally chosen for L3 are "cosine" shapes. Any continuous function may be matched over a 180-degree interval by a Fourier series of cosines only. By making the 180-degree interval the frequency range from 0 to 8.5 mc the cosines are cosines of frequency and can match any gain characteristic if enough terms are used. These cosine equalizers have the following advantages.

1. The range required for any term is less than the total shape to be matched, usually less than half.
2. The residues are always higher harmonics and therefore easily matched if necessary.

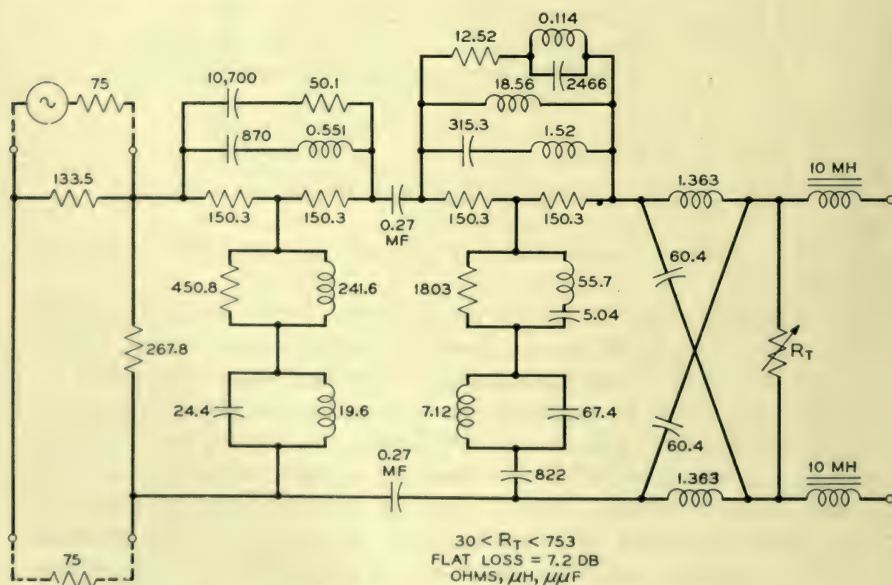


Fig. 7 — Circuit of the regulating network for tube aging.

3. The controls are easily adjusted using methods to be described.
4. If better equalization is required at any point the existing equalizer is retained, additional harmonic terms are added.
5. The major portions of the equalizers require only one value of inductance and two values of capacitance to form the delay lines used in the networks. This also assists in the application of distribution requirements.
6. The manual equalization can be designed with a minimum of information about the system characteristics.
7. The equalization is on a least square error basis rather than minimum peak error.

The networks used to realize these cosine shapes are constant resistance Bode regulating networks³ employing second degree all-pass sections.⁵ If the phase of the all-pass sections were made proportional to frequency, the transmission performance within the frequency band of interest would be that provided by a Fourier series composed of cosine terms in the variable ω . By appropriate choice of the all-pass sections the frequency-phase relationship can be warped to give greater weighting to a specified portion of the frequency range. The phase of the all-pass section is given by

$$\psi = 2 \cot^{-1} \left[\frac{b}{2} \left(\frac{fc}{f} - \frac{f}{fc} \right) \right]. \quad (21)$$

Small b and high f_c weights the low frequencies, the linear phase case $b = 1.2$, $f_c = 10.2$ mc gives uniform weighting and large b weights the high frequencies. A $b = 2$, $f_c = 13.75$ mc, was selected for the L3 equalizers because it weights somewhat the higher frequencies where the television signal is transmitted and second, each unbalanced bridge T network section can be constructed with only four elements, two like inductors and two capacitors. For b 's smaller than 2, coupling between the two like coils is required, and for b 's larger than 2, an additional element is required.

The all-pass networks are designed on a 75-ohm impedance level and thus the flat, normal setting of each regulating network is 4.18 db maximum. The 75-ohm level makes the series and shunt networks identical and also facilitates manufacturing testing. A special dual variable resistor is used as the control element. It has a resistance range of 15 to 375 ohms and provides a regulation range of ± 2.78 db maximum.

A schematic of this network is shown on Fig. 8. In the case of the O harmonic, flat gain, the phase sections are omitted. For the n th harmonic

term, n sections are used in both the series and shunt arms. The constant resistance structure permits the complete equalizer to be formed by a cascade of such networks without interaction effects. This provides a loss characteristic given by

$$\text{Loss} = k_0 + k_1 \cos 2\psi + k_2 \cos 4\psi + k_3 \cos 6\psi + \dots \quad (22)$$

where ψ is the phase of the individual all-pass section which goes from 0 to 90 degrees between 0 and 8.5 mc. The k 's are adjusted by means of the dual adjustable resistors. Note that each term of the series is represented by a corresponding equalizer.

Application of this mop-up polynomial to a large number of L3 deviation characteristics indicates that considerably less range than the maximum ± 2.78 db will be required for most of the cosine terms. Fig. 9 illustrates this convergence of the series for the eight largest amplifier manufacturing variations. This and other studies show that after the first three harmonics the range may be reduced. It can be shown,⁶ for example, that if the system gain deviations are finite within the range of interest (interval of convergence) the coefficients of the approximating polynomial will decrease in magnitude linearly proportional to the number of the terms, that is, the n th coefficient will be smaller than some constant divided by n . If the deviation characteristics are continuous and hence have finite first derivatives, then the coefficients of the approximating polynomial will decrease as the square of

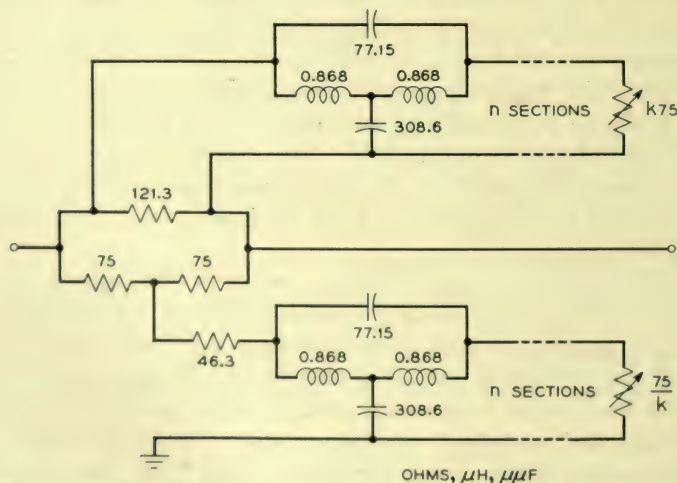


Fig. 8 — Circuit of a cosine equalizer without dissipation correction.

the number of the term. If all derivatives are finite the coefficients will decrease exponentially. This last case is believed to describe the convergence for at least most of the L3 system shapes. Thus for the high order terms that require little range it is possible to reduce the flat loss of the networks.

In order to reduce the flat loss without changes in the 75-ohm impedance level of the phase sections or in the dual adjustable resistors, pads are inserted between the resistance T and the all-pass sections. In this manner the loss could be reduced to 2.2 db for ± 0.5 db range if it were not for the dissipation in the all-pass sections. This dissipation is due to the coils and increases with frequency. It tends to produce a reduced cosine amplitude in the high frequency part of the band. To correct this effect, the pad mentioned above is actually made an equalizer section whose loss change with frequency corrects for the dissipation in the coils thereby yielding cosine amplitudes independent of frequency. The price of this is an increase of the flat loss to 3.4 db for ± 0.5 db range. The circuit of the term 10 network is shown on Fig. 10. The range

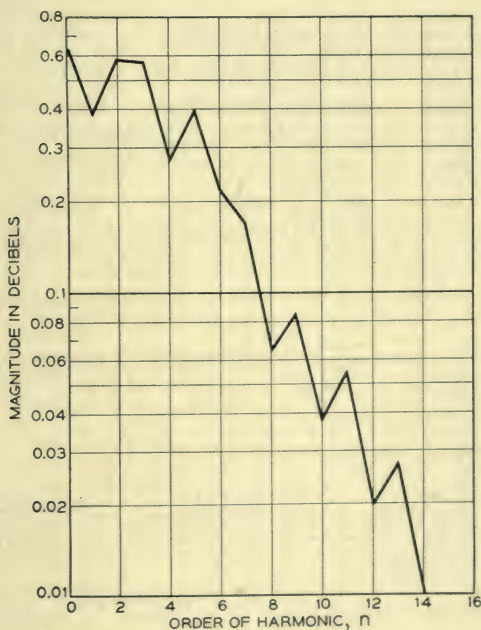


Fig. 9 — Range required for various cosine terms due to expected manufacturing variations of eight critical elements. The magnitudes are based on RSS addition in 25 repeaters.

of this term is ± 0.5 db as is that of all higher terms (to 24). Terms 1, 2, 3 have full range, 4, 5, 6 have 1.5 db and 7, 8, 9 have 1 db. The 0 term, flat gain, is also 1 db because additional flat shape is obtainable from the flat amplifiers used to make up for the equalizer loss.

The shapes provided by the first three terms are shown on Fig. 11. Note that the shapes are cosines of a warped frequency variable and that on this warped scale the shapes are orthogonal. In all, 24 such harmonics plus flat gain are used in the high frequency line for combined systems. For all telephone use only 14 harmonics plus flat gain are required. Fig. 12 shows the construction of one of the cosine networks. These are mounted in groups of five as indicated in Fig. 13. The fixed equalizer and regulating networks are mounted to the rear of the cosine assembly.

Harmonic Adjusting Set

To make a mental harmonic analysis of a complicated gain characteristic is difficult if not impossible. Therefore a special cosine-equalizer adjusting set has been developed which eliminates trial and error from the adjustment process and which leads to a unique optimum adjustment. Broadly the method consists of using sweep frequency methods to convert the gain-frequency characteristic into a repetitive voltage-time function. Gain cosines on the warped frequency scale are converted to voltage cosines of time and the audio harmonic-spectrum components

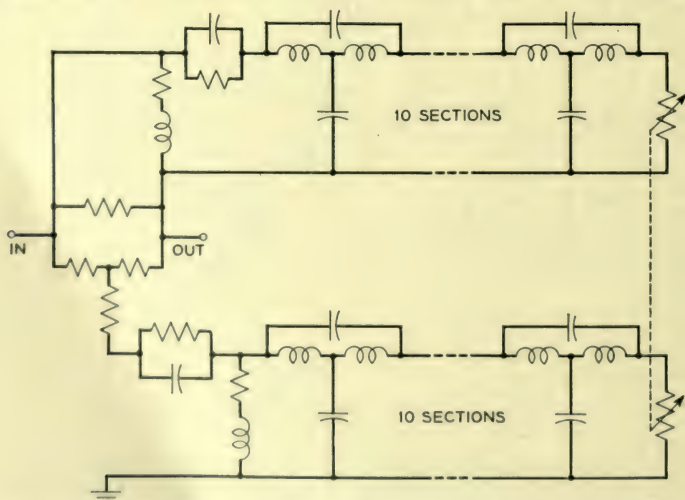


Fig. 10 — Cosine equalizer circuit (tenth harmonic) showing dissipation and range corrections.

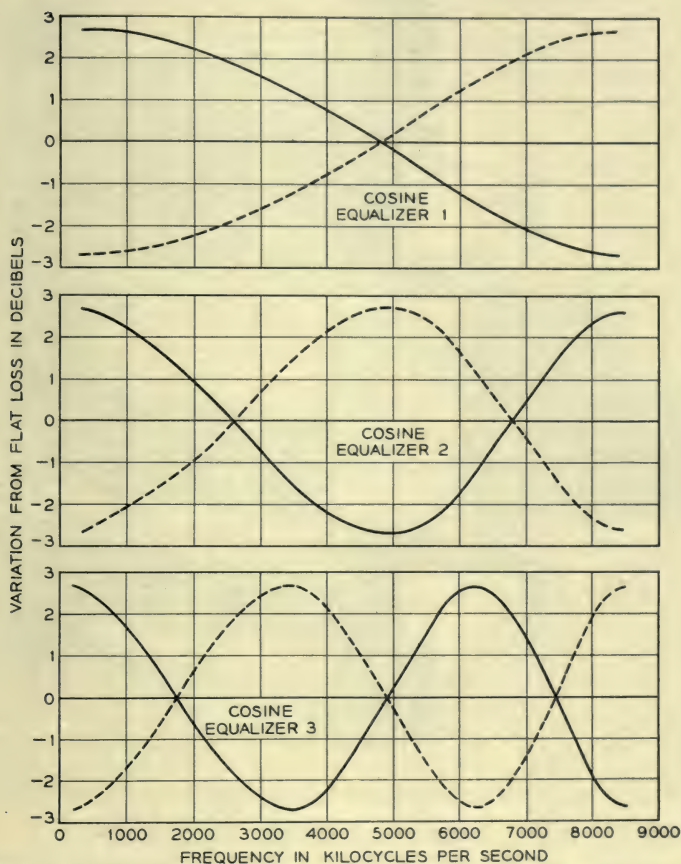


Fig. 11 — Shapes introduced by the first three cosine harmonics.

are individual measures of equalizer control-setting errors. By the adjustment process these audio harmonics are removed thus yielding a gain characteristic describable in terms of only the higher cosine components not available to the equalization operator.

The operation can be explained using the block diagram shown on Fig. 14. The sweep oscillator sends a constant level, variable frequency, over the line and through the cosine equalizer to the detector. The output of the detector on terminals x-x at any instant is a measure of the transmission of the line and equalizer at the frequency being sent by the sweep oscillator at that instant. The sweep frequency starts at zero and sweeps to 8.5 mc in a period t_1 . As shown on Fig. 15 the frequency-time relationship is warped to correct for the warping of the equalizer

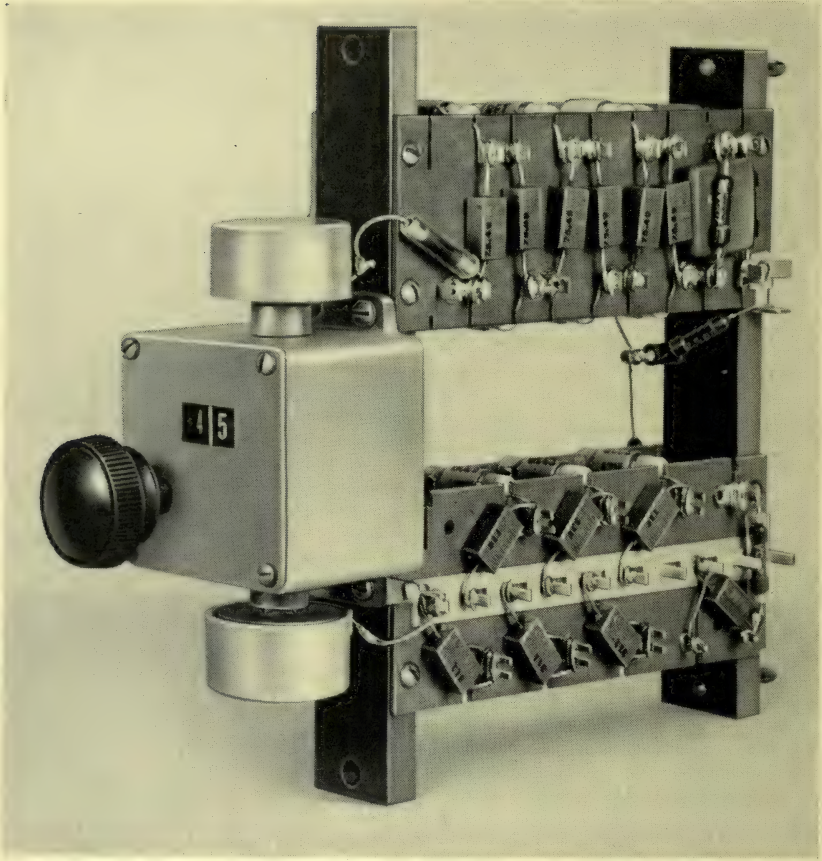


Fig. 12 — View of sing'e cosine term network.



Fig. 13 — Assembly of five cosine terms. A "C" equalizer requires five of these assemblies, a "B" equalizer, three, and an "A", two.

phase-frequency relationship. The sweep oscillator therefore scans at a linear rate in cosine degrees vs time. Upon reaching 8.5 mc the sweep reverses and returns to zero. If the cosine shapes were linear cosines of frequency the sweep would be a triangular wave.

Assume that the line is perfectly equalized except that the first harmonic term is misadjusted. As the sweep goes from 0 to 8.5 mc the voltage at x-x follows the first harmonic curve of Fig. 16 from 0 to t_1 . When the oscillator scans back to 0 the voltage at x-x follows the first harmonic curve from t_1 to $2t_1$. Then the cycle repeats. The voltage at x-x thus becomes a pure cosine of time oscillation of frequency, $\frac{1}{2}t_1$. Although the equalizer shapes are actually cosine on a decibel rather than an amplitude basis this has little practical effect because for small deviations the two are nearly identical.

If instead of a first harmonic error the second or third cosine harmonic

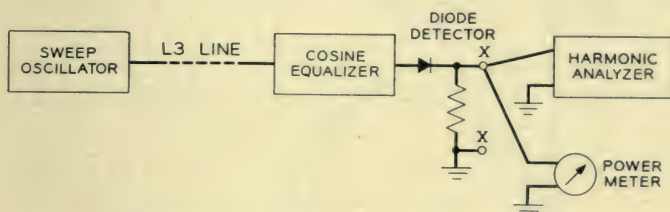


Fig. 14 — Block diagram showing the method used to adjust cosine or other orthogonal equalizers.

is misadjusted, the voltage at x-x will follow the appropriate curve of Fig. 16. The dc component measures the zero harmonic but in practice only the ac components are measured and the flat gain is set as a final step to make the pilot levels correct at the line output. If it takes 0.01 seconds for the oscillator to scan up and back, the first harmonic produces 100 cps output from the detector. The second harmonic equalizer produces 200 cps, the third 300 cps, etc. Therefore at the output of the detector there exist a set of audio frequency harmonics whose amplitudes are a measure of the equalization error of the setting of the cosine controls of corresponding periodicity.

These harmonics can be separated by convention filtering techniques, for example, by an audio tuned detector or harmonic analyzer as noted on Fig. 14. The analyzer can be tuned to 100 cps and the first harmonic control rotated to remove the 100 cps component. Then tuning to 200 cps the second control is operated, etc. This null method is similar to bridge balancing and may be instrumented to similar high precision. After all of the harmonics corresponding to equalizer controls have

been removed the process is complete and the equalization residue must be composed solely of those terms not provided by the equalizer.

The harmonic analyzer method requires tuning or switching and thus to simplify the method still further the actual field equipment uses a power indicator in place of the analyzer as noted on Fig. 14. Given a spectrum of signals of differing frequencies, removing any one reduces the total power. Therefore the entire spectrum may be applied to a power indicator and the reading reduced by adjusting the various equalizer controls. In practice this process is assisted by filtering out the high harmonics which cannot be equalized. This reduces the total power and increases the ease of reading the meter. While the method has been demonstrated using an ordinary 60 cps wattmeter, an electronic wattmeter is used in the field equipment.

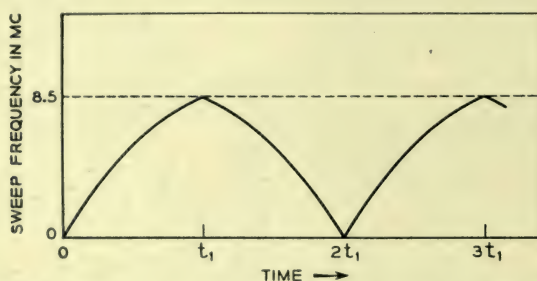


Fig. 15 — Method of scanning the system gain characteristic to convert cosines of frequency into cosines of time.

While the receiver unit can be quite simple, the sweep oscillator is complicated by such things as circuits to control the warping and to hold the sweep limits accurately. In practice the sweep is between 0.3 and 8.5 mc and is at a 37 cps rate. Further there are six pilots on the system and although the dynamic regulators at the adjusting point are paralyzed during the cosine adjustment the pilots to intermediate regulators must not be disturbed. Thus the sweep frequency is shifted very rapidly through the pilots. When the sweep frequency gets within about 25 kc of a pilot it is shifted suddenly to the other side of the pilot frequency. This materially reduces the interference to the pilot without producing transients in the receiver.

While other methods of cosine equalizer adjustment were tested and found to work satisfactorily, the above method was found to be superior. In addition, removal of the filtering permits the power meter to read the rms equalization error and thus the equalization operator can determine the quality of the job and observe whether the state of the line is

satisfactory. Also the power method is usable with sawtooth as well as triangular scanning and, further, works on any set of orthogonal gain or delay shapes. Thus the basic equipment is readily adaptable to the adjustment of D equalizers if their gain and delay shapes are orthogonal.

FIELD PERFORMANCE

In the L3 system the function of the dynamic equalizers is solely to prevent excessive deterioration of the transmission characteristic from one manual line-up to the next. During the manual adjustment the dynamic networks are held at that point in their range which minimizes the probability of running out of range in either direction before the

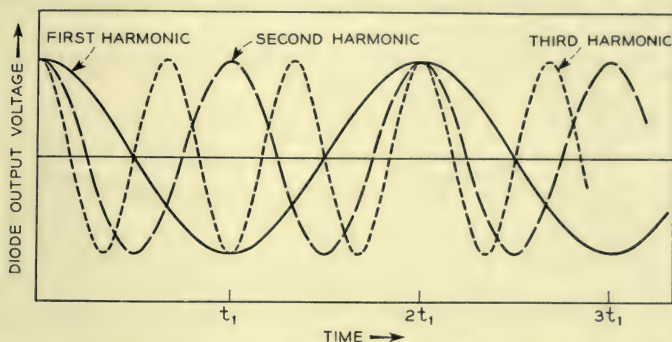


Fig. 16 — Detector output produced by scanning the first three cosine shapes.

next manual line-up. This is done to hold the dynamic ranges to a minimum. Thus the transmission errors remaining at the completion of a manual line-up are chiefly due to the fixed and manual equalizers. It should be noted however that, when the dynamic regulators are restored to operation after the manual adjustment, any residues will be seized by the dynamics and regulated.

As yet the field experience with the regulation system and the dynamic shape performance is quite limited. However, the stability of the regulation system and the action of the computer have been well established. It would appear that the major remaining regulation system problem will be the determination of the cause shapes to the accuracy required for long television systems.

Somewhat more experience has been gained with the cosine equalizers. Using a fixed equalizer design based on ten line amplifiers from initial production, a 100-mile circuit equalized using 15-cosine terms yields the residues plotted on Fig. 17. The ripple at the extreme high end of

the band is largely due to the failure of the fixed equalizer to match the very sharp cut-off of the line between 8.35 and 8.50 mc. Since telephone channels are not transmitted in this region and since television requirements are less severe at such high video frequencies (4.2 mc) this effect does not limit the transmission quality. Also note that long television circuits, over 400 miles, will have a further level of equalization, D. There is no visible impairment of ordinary television pictures due to insertion of the 200-mile L3 line in their path. With critical types of test patterns there is a slight effect. While there are problems yet to be solved before 4,000-mile transmission can be obtained, the 200-mile performance is most encouraging.

In the past the adjustment of manual equalizers has often been a difficult and time consuming task. The cosine adjusting set described

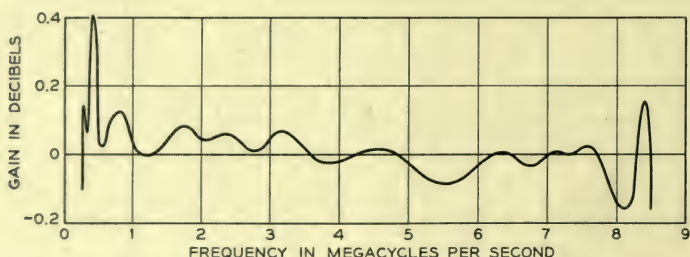


Fig. 17 — Final gain characteristic of a 100 mile L3 line after equalization with 15 cosine shapes.

previously appears to have made sizable inroads on this problem. The adjustment of the 25 controls used for television is a three minute job. The fact that the equalization operator is working toward a unique solution and therefore knows when he is done appears to be of material value.

REGULATION

The L3 regulation system is in many respects, similar to the L1 system. However, the design of a stable regulation system for over five hundred regulators in tandem introduces unusual problems. Also the accuracy and stability requirements have led to the use of novel regulator circuits including, for example, an analog computer as an element of the system.

Six pilots are used, 308, 556, 2,064, 3,096, 7,266 and 8,320 kc. These frequencies were selected to best measure the anticipated system changes as restricted by where signal allocations would permit their insertion.

For example, the wide gap from 3,096 to 7,266 is largely due to the difficulty of inserting and removing pilots in the lower video frequencies of television signals. The problem of finding a satisfactory set of pilot levels and frequencies which will at the same time, be compatible with the desired signals is an important part of the system design problem.¹

The change of four-mile cable loss with temperature is so large (± 1.2 db) that regulation is required at each repeater. It takes three months or more for the cable loss to change 2 db but the normal line maintenance interval is of this order. A gain error of this magnitude could not be allowed to accumulate over very many repeaters before the signal to noise performance of the system would collapse. Other effects such as vacuum tube aging and repeater temperature changes can be allowed to accumulate over as many as 30 repeaters before regulation. These facts dictate the location of regulators in the system. At each repeater there is a "line" regulator controlling a square-root-of-frequency-shape regulating network. Then at equalizing points and dropping points "office" regulators correct for the remaining effects.

CHAIN ACTION

Pilot controlled dynamic regulators derive much of their advantage from the fact that they prevent gain changes from accumulating from repeater to repeater. This advantage is one manifestation of what might be called the "chain action" of a series of regulators. There is however a corresponding disadvantage, disturbances of the pilot cause the accumulation of unwanted gain fluctuations. In previous systems this disadvantage has been aggravated by positive envelope feedback ($1-\mu\beta$ less than one), at some frequencies, an effect known as "gain enhancement". In the L3 system the "gain enhancement" is nearly negligible but the television requirements still require careful control of certain types of gain fluctuations.

The advantage noted above can easily be demonstrated by a simple example. Consider a chain of regulators each having 20 db envelope feedback so that pilot level changes are reduced by 10 to 1. Now consider what happens if each cable section changes loss by one db. Table I illustrates the action.

The first regulator inserts a gain change of 0.9 db in response to the 1.0 db input change. The 0.1 db error increases the input change to the second regulator to 1.1 db and it therefore inserts a 0.99 db correction.

The total resultant error of 0.11 db adds to the change at the third regulator input, etc. Simply stated: The error of the first regulator rides through the system forcing the other regulators to make an accurate

TABLE I

Regulator Number	Input Pilot Change	Inserted Correction	Output Pilot Change
	<i>db</i>	<i>db</i>	<i>db</i>
1	1	0.9	0.1
2	1.1	0.99	0.11
3	1.11	0.999	0.111
4	1.111	0.9999	0.1111

correction. Actually, of course, the above statement is oversimplified but it should be clear that the effective feedback of the regulation system is the (voltage) sum of the feedbacks of the individual regulators. Thus 100 regulators each having 20 db of feedback tend to act like a single regulator having 60 db of feedback. The rigorous treatment of these effects will be developed later.

Fig. 18 shows a block diagram of a regulator in a form intended to indicate the feedback structure. The feedback loop includes a pilot pickoff filter, amplifier and rectifier. This converts the output pilot level into a dc voltage. The “battery”, which is the actual input signal for the circuit, represents the equivalent of the desired pilot output level. The signal applied to the dc amplifier is a dc signal representing the error in pilot level. This dc signal is, in effect, converted back to a pilot level by the action of the regulating network and its modulation of the input pilot level. Thus changes in input pilot level are equivalent to gain changes in the μ circuit of the feedback structure and are resisted by feedback action just as in any other feedback “amplifier”. It is also valuable to note the respective μ and β roles played by the various components since the stability requirements, etc. then become clear. For

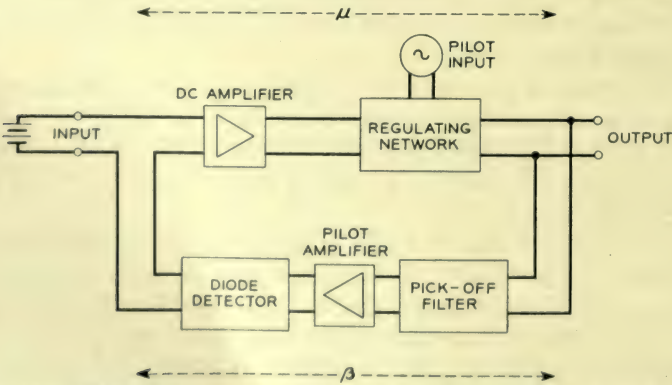


Fig. 18 — Block diagram of regulator showing the feedback structure.

example, the pilot amplifier is in the beta circuit and therefore must be a highly stable device. On the other hand, the dc amplifier is in the μ circuit and its drifts are reduced by the loop feedback.

Having developed the feedback nature of the structure and the roles of the components, the conventional feedback art can be used for the analysis of the individual regulator. One can show that:

$$\frac{\text{Change in output pilot level}}{\text{Change in input pilot level}} = \frac{1}{1 - \mu\beta} \quad (23)$$

$$\frac{\text{System gain change to pilot frequency}}{\text{Change in input pilot level}} = \frac{\mu\beta}{1 - \mu\beta} \quad (24)$$

This result is not very surprising but it can be used to determine the performance of the following regulators in a chain. Many different cases must be considered. Sometimes the pilot levels change because of an effect distributed all along the system. In other cases the change occurs only at the input to the line. Sometimes the pilot level changes are the important effect. In other cases the importance resides in the gain change to the signals. In all cases the results may be complicated by the fact that $\mu\beta$ is, in general, a complex number and thus phase as well as amplitude is important.

Adopting the notation:

ΔP_{in} = fractional change in input pilot at n th regulator,

ΔP_{on} = fractional change in output pilot at n th regulator,

ΔG_n = gain change to signals (near pilot frequency) of n th regulator,
and

ΔG_t = total system gain change = $\sum G_n$,

one can readily show the following:

Case I — Disturbance of pilot only at input to system:

$$\frac{\Delta P_{on}}{\Delta P_{il}} = \left(\frac{1}{1 - \mu\beta} \right)^n \quad (25)$$

$$\frac{\Delta G_n}{\Delta P_{il}} = \mu\beta \left(\frac{1}{1 - \mu\beta} \right)^n \quad (26)$$

$$\frac{\Delta G_t}{\Delta P_{il}} = \left(\frac{1}{1 - \mu\beta} \right)^n - 1 \quad (27)$$

Case 2 — Equal gain change in each regulating section.

ΔP_c = fractional gain change of section

$$\frac{\Delta P_{on}}{\Delta P_c} = \frac{1}{\mu\beta} \left[\left(\frac{1}{1 - \mu\beta} \right)^n - 1 \right] \quad (28)$$

$$\frac{\Delta G_n}{\Delta P_c} = \left[\left(\frac{1}{1 - \mu\beta} \right)^n - 1 \right] \quad (29)$$

$$\frac{\Delta G_T}{\Delta P_c} = \left[\frac{1}{\mu\beta} \left(\frac{1}{1 - \mu\beta} \right)^n - 1 \right] - n \quad (30)$$

As an example of the application of these formulas consider the effect of television induced compression. The presence of the television signal reduces the gain of the line amplifier. The effect is small but cumulative. In the absence of regulator action it merely compresses the television signal slightly and makes a negligible change in the contrast rendering of the picture. However the regulators observe a gain change to the pilots and attempt a correction. The very rapid changes are ignored but 60 cps, for example, is partially corrected. This introduces a 60 cps gain change which will lag the picture and therefore must meet 60 cps bar pattern requirements. This problem is solved by keeping the regulator response low at 60 cps.

For $\mu\beta$ of -70 db, 90 degrees, at 60 cps a chain of 700 regulators will insert a total gain change approximately one tenth that of the total compression. If the $\mu\beta$ were allowed to approach -50 db the total gain change would equal the compression and certain types of pictures would be degraded.

The above example brings out one of the important facts: When designing regulators for long systems the $\mu\beta$ characteristic must be carefully controlled to losses much higher than is customary in amplifier design. In a conventional feed-back-amplifier loop-cutoff the magnitude and phase of $\mu\beta$ is no longer of much interest after it drops below -10 db. In L3 regulators the loop is of vital interest to losses of the order of 70 db. This is, of course, largely due to the fact that the chain action increases the effective system feedback by nearly 60 db. Thus the over-all system is similar to conventional amplifier practice. Loop gain ($\mu\beta$) and feedback ($1-\mu\beta$) characteristics for the line and office regulators are shown on Figs. 19, 20, 21 and 22. Note that 1,000 line regulators in tandem give an over-all gain enhancement of only 1.2 db. This would be even less if it were not for a 100 cps roll-off in the dc amplifier to reduce noise. One hundred office regulators give 0.8 db gain enhancement even with their 20 cps roll-off.

THE DYNAMIC LINE REGULATOR

As indicated in Fig. 23 a crystal filter is used to pick the pilot off the line in the presence of the other signals. The filter impedance goes

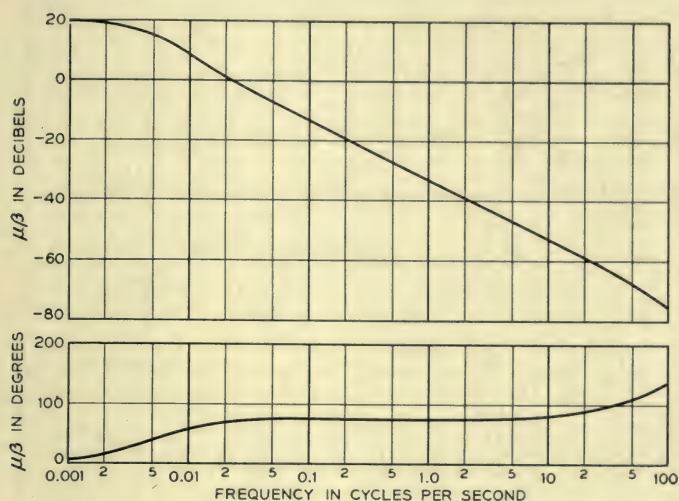


Fig. 19 — Loop gain characteristic for the line regulator.

through resonances due to the crystals and introduces a gain characteristic on the line that cannot, in practice, be equalized. Thus it is necessary to hide the filter from the line with loss. To hold the transmission distortion of signals to 0.15 db with 500 regulators (0.0003 db per regulator) requires a voltage loss of 23 db with the filter impedance changing by large factors from its nominal 25,000-ohm level. The power loss bridging on the 37.5-ohm (75 into 75) circuit is, of course, much greater.

The 7,266-kc pilot level at the line amplifier output is -16 dbm into 75 ohms. The filter and pad loss totals 24 db (voltage ratio) leaving an

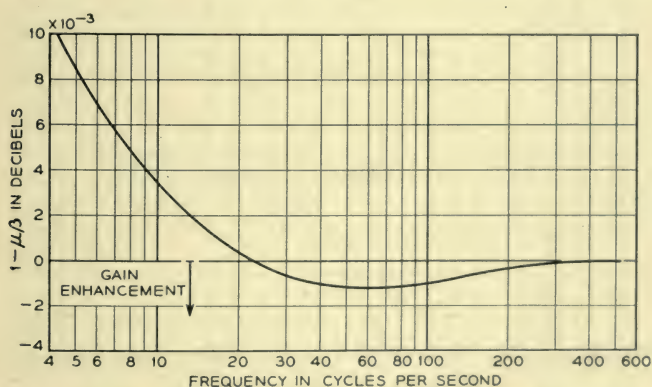


Fig. 20 — Feedback characteristic for the line regulator showing the gain enhancement effect.

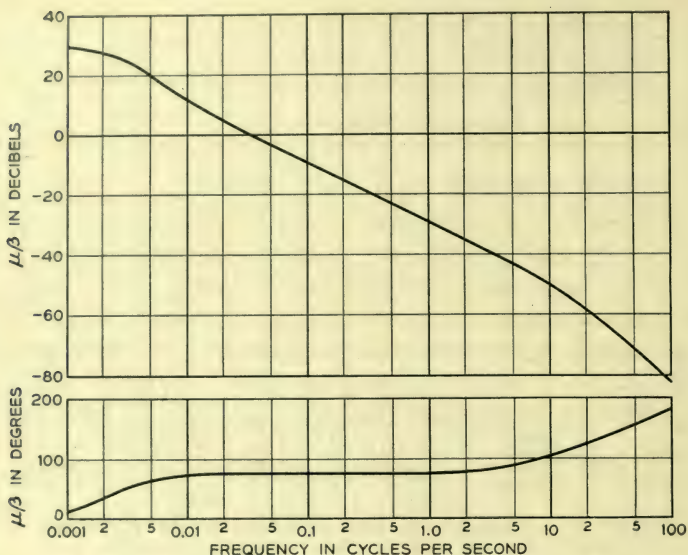


Fig. 21 — Loop gain characteristic for office regulators.

available pilot signal of about 0.002 volts in 25,000 ohms. In order to solve drift and stability problems in the dc circuits the pilot is converted to a dc voltage of 60 volts. This requires an amplifier-rectifier of 90 db voltage gain, stable with time and temperature.

As indicated on Fig. 23, the amplifier consists of three stages using

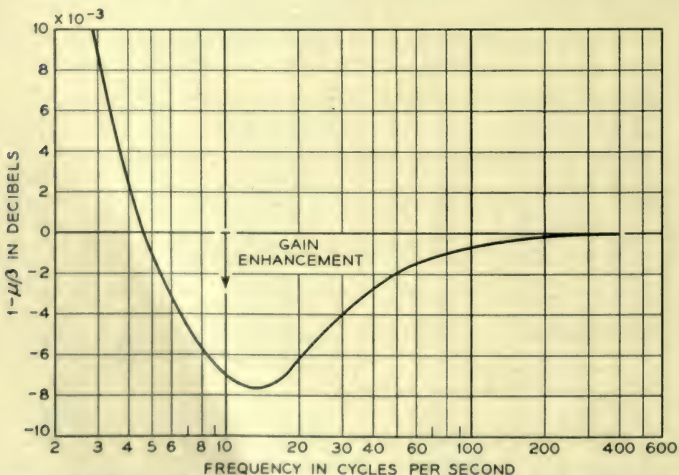


Fig. 22 — Feedback characteristic for office regulators showing the gain enhancement effect.

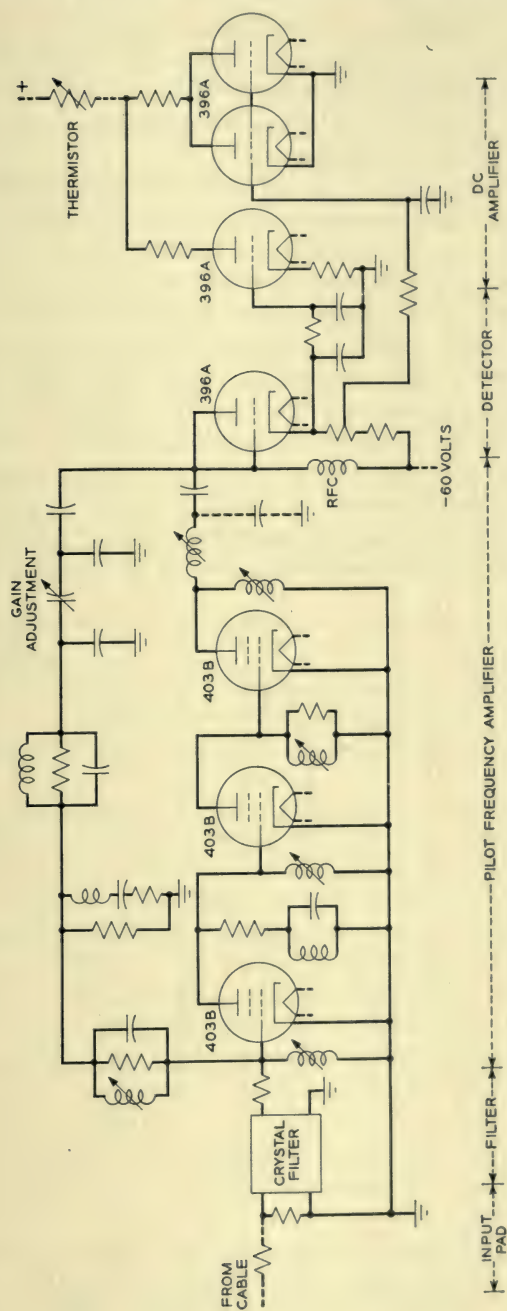


Fig. 23 — Simplified schematic of the line dynamic regulator showing the signal circuits.

403B (long life 6AK5) tubes. Feedback is taken shunt-shunt, output plate to input grid, to stabilize the input and output tuned circuits against Q changes with temperature. The regulators are designed to operate from -20°F to $+160^{\circ}\text{F}$ and the amplifier gain does not change by more than 0.3 db over this range.

The beta circuit contains an adjustable condenser divider for field gain adjustment. The tuned sections are heavily damped to get temperature stability but nevertheless compensate for the cutoffs of the μ circuit. The resultant loop gain and phase are shown in Fig. 24. It will be noted that the phase margin against singing is rather large, about 60 degrees. This permits tube replacement without retuning as well as protection against tuning changes due to time and temperature.

Grid-plate capacity places a limit on the permissible interstage impedance for stability. Thus the output circuit between the third stage and the rectifier is operated as a reactive transformer giving a voltage step-up of 2. This increases the output voltage obtainable without raising the impedance facing the third tube above 12,500 ohms.

About 16 db of local dc feedback is used on each stage to stabilize the

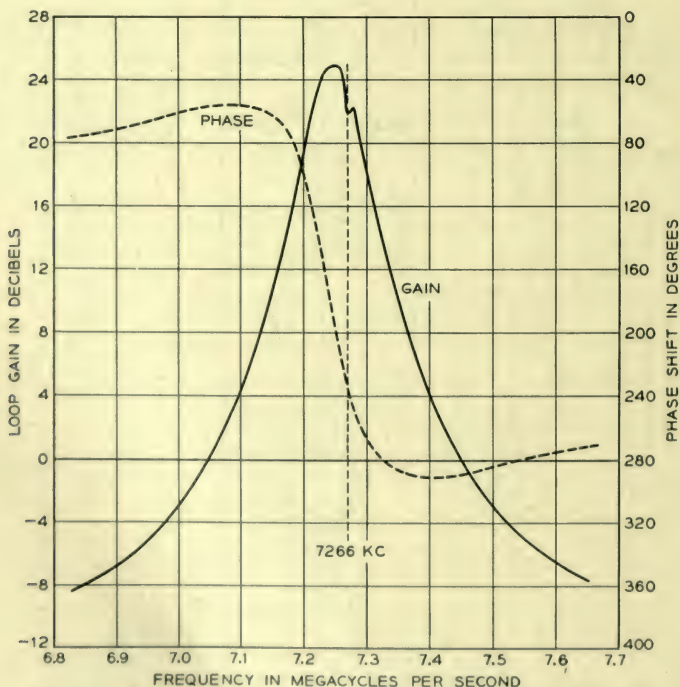


Fig. 24 — Loop gain and phase of the 7266 kc pilot amplifier.

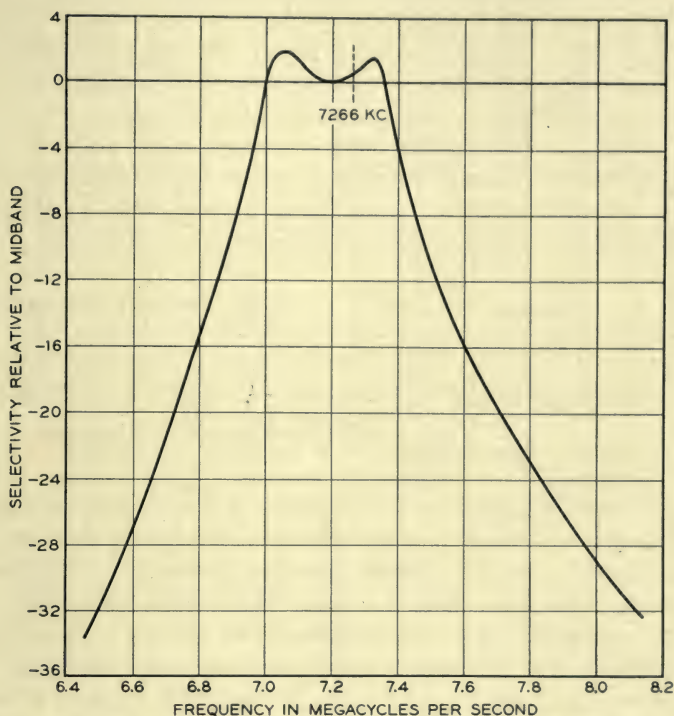


Fig. 25 — Regulator selectivity without the crystal filter.

cathode current. In so far as trans-conductance depends upon cathode current, trans-conductance changes are reduced. This is of material value in further reducing the effects of vacuum tube aging.

Fig. 25 shows the external gain of the 7,266-kc amplifier without the crystal filter. The filter response is shown in Fig. 26. The relatively wide 3 db bandwidth of ± 1.5 kc is to reduce the contribution of the filter to the gain enhancement problem. The large rejections to frequencies further removed from the pilot prevents operation of the regulator by signals other than the pilot. Also note that for very strong signals such as the television carrier at 4,139 kc the filter is aided by the amplifier selectivity.

The diode detector is also designed to reduce the effects of interference. The time constant of the detector is made short so that the output will follow envelope fluctuations up to about 40 kc. Thus the output of the detector in the presence of an interfering signal within 40 kc of the pilot becomes the power sum rather than the voltage sum of the two signals. This is readily understood from Fig. 27. Here E_p is

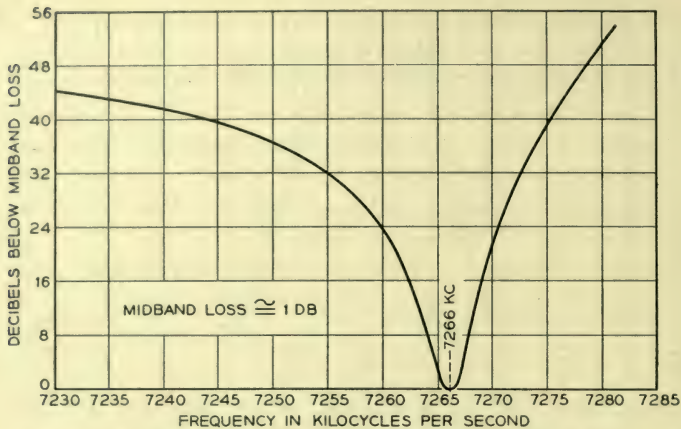


Fig. 26 — Loss characteristic of the 7266 kc crystal filter.

the normal rectified pilot. In the presence of the interfering signal E_i , a diode detector with a long time constants will deliver a dc output of $E_p + E_i$ and thus give voltage addition between the pilot and the interference. With a fast time constant the detector output can follow the nearly sinusoidal envelope variations and the dc level is changed only slightly. The ac component may be suppressed by the cutoff of the dc amplifier, but, even if this is not the case, the thermistor being a thermal device responds to the total power rather than the peak amplitude. Thus the diode time constant is made long enough to hold over a few cycles of the pilot frequency but short enough to follow the important interference difference-frequencies.

The dc amplifier consists of three triode sections essentially in parallel

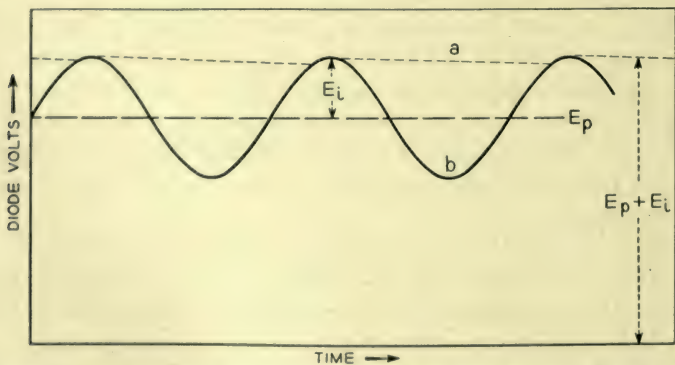


Fig. 27 — Diode detector output in presence of interference E_i . E_p is normal pilot signal. Curve a obtains with long time constant, b with short time constant.

but with the biases and feedbacks differing in order to provide an EI characteristic that corrects for sensitivity changes of the thermistor with operating current. This maintains the overall loop feedback relatively constant over the 1 to 20-ma current range. The thermistor is directly heated by the plate current of the dc amplifier in order to obtain single time-constant performance of the thermistor. The thermistor transmission, plate current changes as an input and pilot level as an output, is the main frequency characteristic of the regulator loop.

LINE THERMOMETER REGULATORS

It is possible to dilute the regulation system with less costly, less accurate regulators without undue loss of overall performance. This is

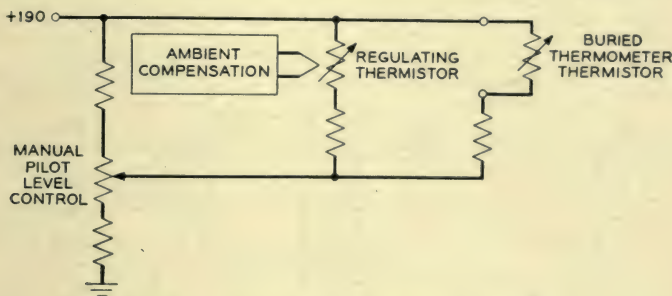


Fig. 28 — Thermometer regulator schematic.

accomplished by the use of thermometer regulators at alternate regulating points. These consist of a thermometer thermistor buried in the ground, electrically in parallel with the regulating thermistor. The circuit is quite simple as indicated by Fig. 28. Ground temperature changes vary the resistance of the thermometer thermistor thereby changing the current and resistance of the regulating thermistor. The manual control is used to effect initial alignment of the system. The regulating sensitivity is designed to slightly overcompensate for cable loss changes in order to somewhat ease the burden on the following dynamic regulator.

AMBIENT TEMPERATURE COMPENSATION

Both types of line regulators require the assistance of ambient temperature compensation of the regulating thermistor. Conventional compensation circuits would hold the thermistor resistance within about 20 per cent but this would produce an error of one db at a thermometer

regulator and about 0.2 db at a dynamic unit. Thus an improved compensation scheme was required which would connect only to an indirect heater, the bead itself being already controlled by dc heating from the regulators.

The compensation circuit adopted is shown on Fig. 29. A second thermistor called the compensating thermistor is mounted in the same glass envelope with the regulating thermistor. The fixed resistor R_1 and the compensating thermistor together with transformer T form a bridge which is made a feedback path for tuned amplifier A . The feedback is positive when the compensating unit is cold so oscillation begins at the tuning frequency (4 kc). These oscillations heat the thermistor and tend to bring the bridge into balance. The bridge stabilizes at a

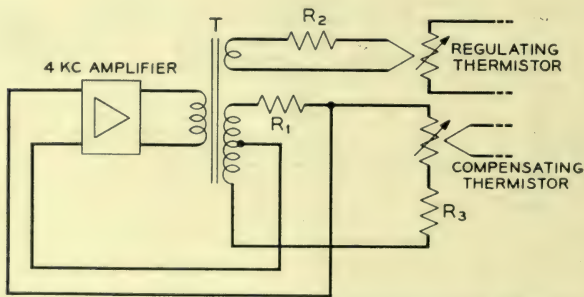


Fig. 29 — Ambient temperature compensation oscillator.

small unbalance just sufficient to yield a loop gain of unity. The level of oscillation is forced to that value which will maintain this small unbalance. Any changes in balance thereafter produce deviations of loop gain from unity and the oscillation level increases or decreases until the equilibrium is reestablished. Thus the level of oscillation changes with temperature but the resistance of the compensating thermistor is held constant.

Because the circuit supplies nearly perfect temperature compensation to the unit in the bridge a suitable fraction of the oscillator power may be fed to the heater of the regulating thermistor to achieve very close compensation of it. Resistors R_2 and R_3 are adjusted in manufacture to correct for slight differences between the two thermistors. Note that the compensating thermistor is provided with an unused heater to match the thermal properties of the two units.

The amplifier consists of a single 403B tube with 20 db dc feedback for current (and transconductance) stabilization. This feedback is vital

because it also introduces a slight compressive action in the amplification of the tube and thereby prevents rapid wild changes in oscillation level. Bypassed dc feedback on an amplifier causes dc second order distortion to increase the bias and thereby reduce the transconductance. This effect overcomes the tendency of the third order distortion to create expansion in this particular tube. If the thermistor response were fast compared to the reciprocal of the bandwidth of the amplifier the compression action would be unnecessary. However with an audio frequency amplifier and a 100 second thermistor the compression is essential in preventing motorboating.

The field limits on the compensation of the regulating thermistor over the range -20 to $+160$ degrees F are ± 3 per cent in resistance due to all causes including manufacture and aging. Specific units can be adjusted to yield compensation to a fraction of a per cent.

OFFICE REGULATORS

The L3 office regulators are similar in design to the line regulator. However the office regulators operate their regulating networks via an analog computer and, of course, a variety of pilot frequencies are employed. Because signals are dropped at offices, higher loop feedbacks are used to insure accurate equalization. However, temperature variations are smaller and conventional thermistor ambient temperature compensation is adequate. Also the smaller number of office regulators permits less isolating loss for nick effect (except for the 7,266-kc office regulator).

The lower levels of the pilots (except 7,266) are compensated by reduced isolation loss, (12 instead of 23 db), and reduced detector level (40 instead of 60 volts). Thus the gain required is not substantially increased. The pilot amplifier design is therefore different primarily in the tuning frequency and in the simplifications in the lower frequency units permitted by the higher permissible interstage impedances.

The diode detector feeds a cathode follower to obtain the dc voltage representing the deviation of the pilot from its assigned value as a low impedance source to feed the computer. The appropriate signals from the computer are fed to the dc amplifier. This amplifier differs from that used in the line regulators in that (1) a push pull input is provided, (2) higher gain is required to produce greater feedback, 30 db, and overcome computer losses, 5 db and (3) the output stage supplies somewhat higher currents, 1 to 30 ma, (except 7,266) because the regulating networks use a lower impedance thermistor.

OTHER REGULATOR FUNCTIONS

The regulators are also used for alarm and pilot indicator functions. At line dynamic regulators the current flowing through the diode detector load resistance is also passed through a relay to obtain an alarm indication whenever the pilot level deviates from normal by more than 3 db. At offices similar arrangements operating on a 2 db error are provided both for alarm and switch initiation purposes. If any pilot deviates by 2 db the service switches to the spare line. In addition fast switch initiation is obtained from the 7,266-kc regulator by direct connection to the detector output. This arrangement avoids the time delay of the relay operation. For pilot level indication the diode load current is read on appropriate meters. This avoids the necessity of providing separate pilot level indicators and is possible because of the reliability and stability of the regulator amplifiers.

MECHANICAL FEATURES

Figs. 30 and 31 are views of a regulator seen from the wiring side and from the top respectively. The chassis consists of two steel end plates riveted to steel angles, with a punched copper plate screwed to this



Fig. 30 — Line dynamic regulator as seen from wiring side.

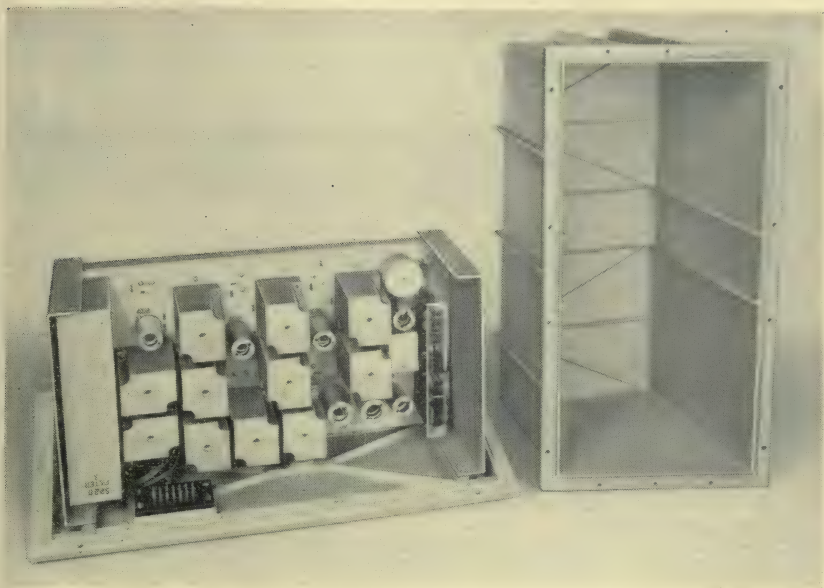


Fig. 31 — Top view of line dynamic regulator showing case.

structure. The salient feature of this type of construction is that one universal punched plate can be used for all the regulators (including the thermometer regulator), any individual regulator being fabricated by mounting the necessary component cans and vacuum tube sockets on it. Power wiring and all leads that are not critical as to length or placement can be run in the wiring trough around the edge of the chassis shown on Fig. 30. This eliminates the necessity of lacing the wires into a cable, a significant saving in production effort.

The component cases are shown on Fig. 31. They are zinc die-castings and contain network elements assembled on stypol forms which fit inside the cans. One universal case accommodates sixty-four different combinations of elements required by the various regulators. The necessary wiring of the individual cases may be completed before assembly on the regulator chassis. This feature also saves production effort.

The whole chassis is mounted inside a die-cast zinc housing, and all power and test leads are brought into the regulator through airtight connections. The two parts of the housing when assembled together are made airtight by a rubber gasket which fits into the slot around their inner edges. The general construction features and size of the regulator assembly can best be understood by inspection of the illustrations.

ACKNOWLEDGEMENTS

Space does not permit listing all of the people who contributed to the success of this work. However, we wish to mention S. A. Levin for his work on cause shapes, R. H. Klie for his leadership in system studies, E. T. Harkless for his study of cosine equalizers, and E. Ley for mechanical design of equalizers. Also mention should be made of C. J. Custer and E. G. Morton for their work on regulator circuits, A. R. Rienstra for studies of gain enhancement, C. H. Bidwell for "chain action" analysis, and F. R. Dickinson for mechanical design of regulators.

REFERENCES

1. C. H. Elmendorf, A. J. Grossman, R. D. Ehrbar, R. H. Klie, The L3 Coaxial System — System Design, see pp. 781 to 832 of this issue.
2. L. H. Morris, G. H. Lovell, F. R. Dickinson, The L3 Coaxial System — Amplifiers, see pp. 879 to 914 of this issue.
3. H. W. Bode, Variable Equalizer, Bell Sys. Tech J., April, 1938.
4. R. L. Lundry, Attenuation and Delay Equalizer for Coaxial Lines, A.I.E.E., Trans., **68**, 1949.
5. P. H. Richardson, Variable Attenuation Network, patent 2,096,028.
6. H. S. Carslaw, Introduction to the Theory of Fourier Series and Integrals, Third Edition, 1930, MacMillan & Co., Page 148.

The L3 Coaxial System

Amplifiers

By L. H. MORRIS, G. H. LOVELL and F. R. DICKINSON

(Manuscript received April 17, 1953)

The line amplifiers for the L3 coaxial system are designed to compensate for the loss of the four miles of cable which separate the repeaters; the flat amplifiers are used to compensate for equalizer loss and as transmitting amplifiers. The two types are basically similar, consisting of two feedback amplifiers in tandem separated by an inter-amplifier network; in the line amplifier this network is variable and is automatically adjusted to compensate for variations in cable temperature and for small deviations from the nominal four-mile spacing.

Coupling networks employing high-precision transformers are used to connect the amplifiers to the coaxial cable through the required power separation filters. The low impedance windings of the transformers are center-tapped and a balancing network provided in order to match the cable impedance over the transmitted frequency band. The amplifiers are equipped with plug-in tubes of high figure of merit which were developed for this application. A double-triode output stage is used to obtain improved system signal-to-noise performance. Provision is made for preventive maintenance of vacuum tubes and for a controlled adjustment of gain on an in-service basis.

All important components of the amplifier are subject to quality control procedures to assure that the average gain of groups of amplifiers will be held within narrow limits and that individual amplifiers will form a normal distribution around the average. This approach is essential in order to meet system equalization and signal-to-noise objectives. Careful mechanical design and rigid control of the mechanical aspects of manufacture are necessary to minimize gain variations which might be caused by variations of parasitic circuit elements and unwanted feedback effects. Special measures were required to keep the temperature rise within the sealed die-cast housing within tolerable values.

INTRODUCTION

In a transmission system such as the L3 coaxial, the degree to which system objectives are achieved is largely dependent on the quality of the amplifiers which compensate for the cable loss. To a considerable extent the same statement applies to the similar flat-gain amplifiers used to make up for the loss of the equalizers at various points along the route. The development of an amplifier which would meet the exacting requirements of the L3 system was in turn dependent on new developments in the fields of vacuum tube design and circuitry, network design techniques, element and network fabrication, and statistical quality control. To these new tools were added the lessons learned in years of manufacture and operation of the preceding L1 system.

The importance of some of these factors can best be illustrated by examining the implications of the amplifier requirements which follow from the material in the companion papers on system design¹ and equalization.² Obviously the figure of merit, modulation coefficients and life of the vacuum tubes will be determining factors in setting the amount of feedback that can be obtained and the signal-to-noise ratio of the system. It is not so immediately apparent that system requirements could not be met with the present 4-mile repeater spacing if it were not for the use of quality control at every stage of manufacture from elements, and even the raw materials entering into components, to complete amplifiers. As the companion papers show, however, the equalization plan of the system is predicted on a degree of reproducibility of amplifier gain and delay characteristics obtainable only by quality control applied at every stage of the manufacturing process.

The present equalization plan is based on the assumptions that the gain of the average line amplifier will match the loss of the preceding line section to within 0.15 db, and that the average amplifier gain will not vary from one batch of new amplifiers to another by more than about 0.06 db. Under these assumptions a system equalization plan can be worked out which results in reasonable spacing between equalizers, a tolerable signal-to-noise penalty due to misalignment and equalizer loss, and a practicable procedure for adjusting long systems. Any gross departure from the basic assumptions as to reproducibility of amplifiers would seriously compromise these objectives, which even now are achieved only by using equalization which requires a flat gain amplifier for every four or five line amplifiers. Now it turns out that with the most precise elements that can be made, the gains of individual amplifiers will vary by about ± 0.6 db. We need, therefore, a tool which will permit us to control the gain of the average amplifier to an order of magnitude

greater precision than we can economically control the individual. This is exactly the effect which modern methods of statistical quality control aim at achieving, and since the entire amplifier is merely the sum of its parts, quality control must start at the roots of the manufacturing process. In order to apply quality control intelligently, and to be sure that all important causes of gain variation are understood, it has been necessary to carry out, side by side with empirical laboratory work, a program of computing the insertion gain of the amplifier, starting with fundamental element values. These computations have also proven of value in obtaining satisfactory stability margins in the design of the low and high frequency cut-offs of the feedback loops, and in obtaining preliminary information on amplifier gain deviations for use in equalization planning.

The severe gain and delay reproducibility objectives also have their effect on the mechanical design of the amplifier. At first sight the unit appears to be a lumped constant structure rather than one in which distributed effects would be of paramount importance. Usually the circuit designer in such a case is interested in the mechanical design only for reasons of neatness and economy, but when we look deeper we find that in the amplifier structure as a whole as well as in the case of certain components, the effects of distributed capacity and inductance could easily defeat our objectives if the mechanical design were not such as to assure precise control of element placement and wiring lengths.

For these reasons, an order of mechanical accuracy is specified beyond that which can be justified by the accuracy of transmission measurements on individual amplifiers. This striving for mechanical accuracy is carried all the way through, from element piece parts through subassemblies to the assembly on the final amplifier framework. The logical expectation is that by reproducing the amplifiers as exactly as possible we will control not only the known elements but also the many parasitic effects, and thereby minimize the appearance of shifts in amplifier gain and delay which would be substantial in the system even though difficult to detect in the measurement of individual amplifiers.

Another design feature of the amplifier based on the system equalization point of view is the omission of all adjustable elements, or trimmers, to control the transmission. By eliminating gain adjustments, the possibility of adjusting one element to compensate for the short-comings of another, and the possibility of systematic errors of setting due to faulty or inaccurate adjustment techniques, are both eliminated. Either of these possibilities would tend to convert relatively large random effects into smaller but systematic effects, a conversion which would penalize

the system equalization. Obviously, however, the elimination of gain adjusting elements puts a higher premium than ever on the requirement that all elements, including tube capacities, show only small random variations about tightly controlled nominal values.

In addition to the gain requirements discussed above, the following line amplifier objectives are to be met:

1. The gain of the amplifier must be continuously variable, under the control of a regulator circuit, to compensate for differences in lengths of cable sections and for variations of cable loss caused by temperature changes. The shape of the gain change should match the square root of frequency shape of the line loss change over the transmitted band to within a few hundredths of a db. This accuracy of shape, which of course is based on equalization considerations, should hold over the entire regulation range, which is about ± 6 db at the top transmitted frequency.

2. The input and output impedances of the amplifier should match the cable impedance in order to minimize the effects, on television transmission, of echoes caused by line irregularities such as splices. Tolerable values of reflection coefficient at amplifier input and output are about 5 per cent at the television carrier and 10 to 15 per cent at upper band edge.

3. Feedback consistent with system modulation requirements, shaped across the transmitted band to minimize low frequency intermodulation products and to give a smooth, easily equalized shape of gain change as tubes age, must be obtained while maintaining adequate margins against singing.

4. Other requirements include design and selection of elements to assure as small a change of gain versus ambient temperature as possible, mechanical design provisions for keeping the temperature rise within the unit to a minimum both in order to keep the temperature-gain effect small and to obtain long element life, a sealed housing to avoid damage by humidity in exposed locations, and provision of facilities for testing tubes in service.

CONFIGURATION

The circuit configuration of the line amplifier is shown in Fig. 1. It consists of two independent feedback amplifiers with a regulating network between them acting like a two-terminal interstage. Each of the amplifiers is essentially a two-stage circuit — the input amplifier literally so and the output amplifier essentially so since the double-triode output circuit acts like a single stage. Each amplifier is connected to the coaxial through a coupling network which consists of a transformer

OUTPUT AMPLIFIER

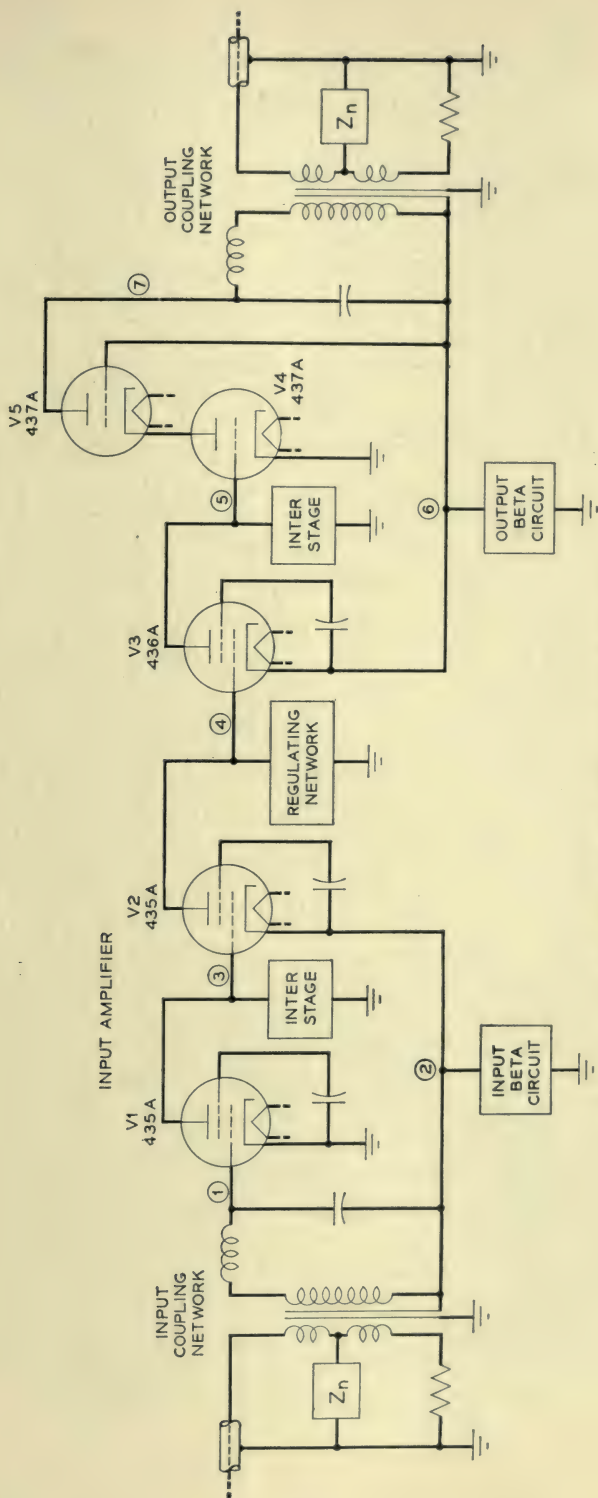


Fig. 1 — L3 line amplifier configuration.

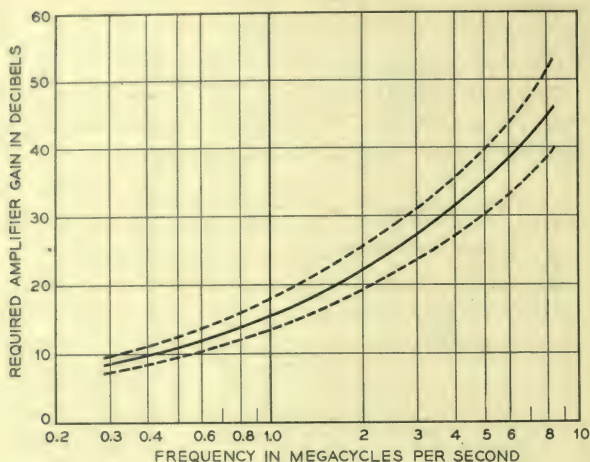


Fig. 2 — Required line amplifier gain.

plus gain shaping and impedance adjusting elements. Since there is no feedback around the coupling networks, they directly affect the insertion gain of the amplifier, as do the two beta circuits and the regulating network. The required shaping of insertion gain across the transmitted band is obtained and controlled by the design of these five networks. Fig. 2 shows the required amplifier gain for nominal and extreme thermistor settings; these required gains differ from the line loss by the small losses of the associated repeater components. At mid-range thermistor setting about 37 db of gain shaping is needed. The manner in which this shaping is distributed among the five networks has important effects on the feedback which can be obtained and the sensitivity of amplifier gain to element variations.

This configuration offers several advantages, one of the most important of which is that the regulating network is between the amplifiers. In most other configurations, the only gain-determining network available for the regulating function is the beta circuit. When the feedback is not infinite, this introduces errors for which it is difficult to compensate, and limits the available feedback by complicating the design. In this configuration, the impedances which the amplifiers effectively present as shunts across the regulating network are high and can be allowed for in the design so that the regulation error can be made much smaller than the error associated with beta circuit regulation in an amplifier having relatively little feedback. Other major advantages are the superior signal-to-noise performance and the relative simplicity of the feedback loops of this amplifier as compared to alternative designs.

MECHANICAL DESIGN

The mechanical assembly of the amplifier, like the configuration, is divided into three main sections. The input and output amplifiers are mounted on separate chassis which are designed for ready removal from the main base casting. The regulating network is mounted in an enclosed, shielded compartment which also serves to shield the component amplifiers from each other. Figs. 3 and 4 show the assembled amplifier.

Because of the wide frequency range and close control of parasitic capacity and lead inductance required, the L3 amplifier was designed as an integrated whole, and all networks were designed with, and as part of, the amplifier. In each case circuit elements were placed in space in the best possible position for optimum electrical performance, and supporting structures were then designed to maintain the desired space relationships. These supporting structures are made as separate units which mount on the amplifier chassis, so that the networks can be individually tested before final assembly into the amplifiers, and can be removed for repair or replacement if necessary.

Heretofore, this method of design has been impracticable because it results in very complicated supporting structures. It was feasible in this case because of the availability of a new type of material, which removes many of the mechanical design constraints. This is a cold casting resin, and parts are produced in a cheap phenolic mold. Since the process is practically equivalent to the sand-casting of metals, complex parts can be economically manufactured even in the relatively small quantities required for L3 production. Where necessary to assure accurate

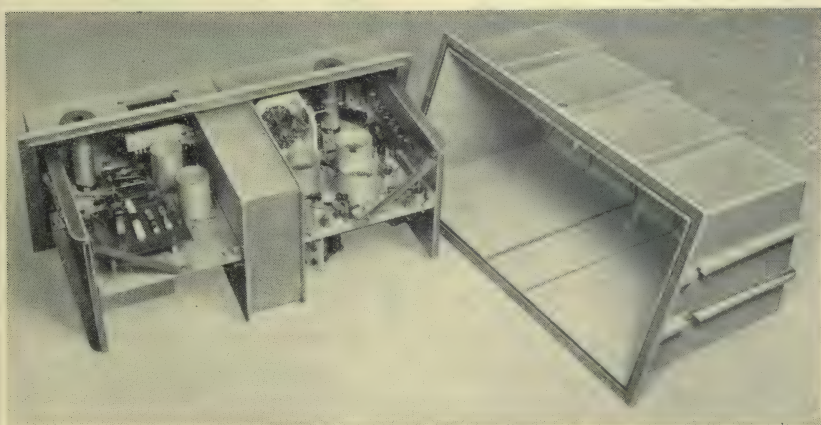


Fig. 3 — Line amplifier and housing.

location or orientation of parts, the cast resin frameworks are milled or spot-faced. The structure used in the input interstage, shown in Fig. 5, is an example. Circuit elements are wired to pins driven into the casting or, in some cases, to wires imbedded in the material. All wiring can thus be made direct with a minimum of bending and no doubling back, resulting in a reproducible and uniform product.

The entire amplifier is housed in a sealed die cast container to protect the components from humidity damage. Dessicant is enclosed in each amplifier. It would have been desirable to make this housing of aluminum, but the high melting point of this metal makes it impractical for such a large casting to meet the air-tightness requirement. A zinc-base die casting alloy was therefore used. Sealing is accomplished by rubber gaskets at all openings for connections and at the joint between the two parts of the housing. The removable part of the casting which serves as a cover is made as large a part of the total housing as possible, in order to provide maximum accessibility for maintenance of the units mounted on the base. The entire housed unit is arranged to mount on a relay rack mounted panel by means of slides which are self-locking. Signal and power connections are made by means of flexible cords which

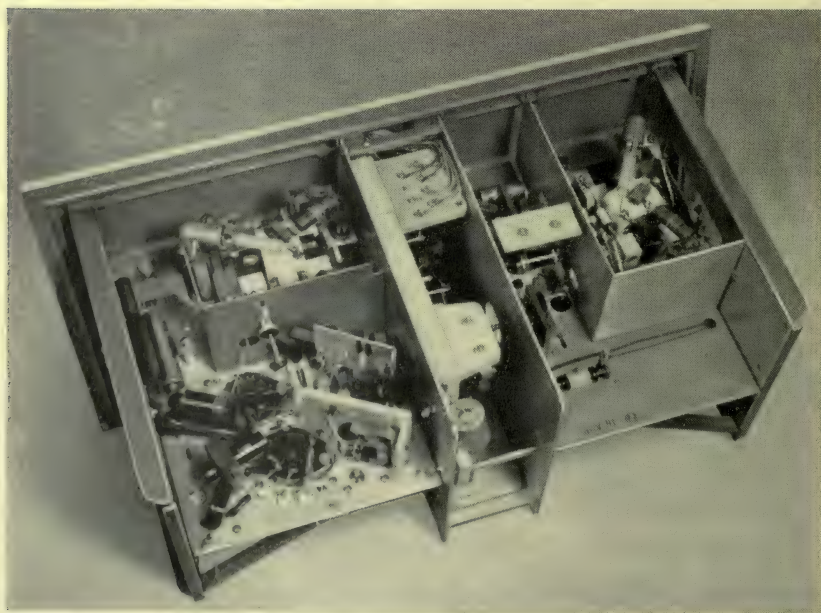


Fig. 4 — Line amplifier, wiring side.

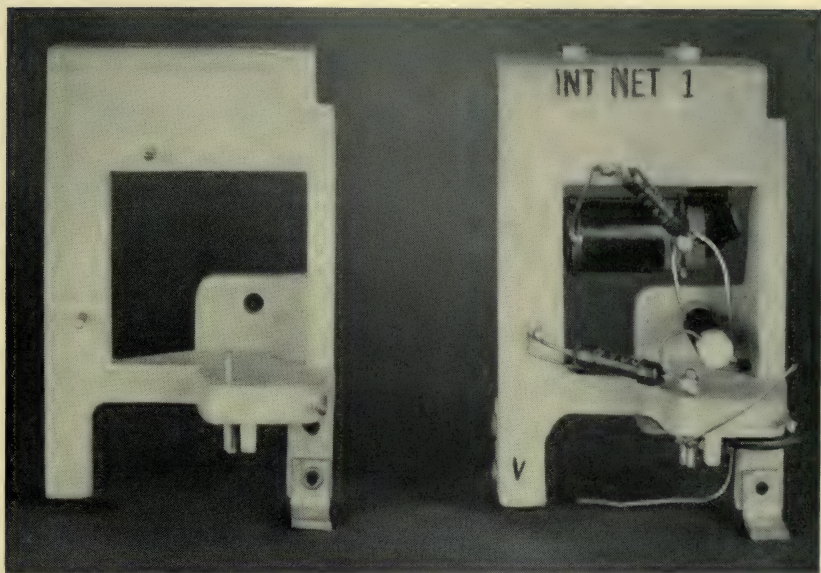


Fig. 5 — Input interstage, illustrating use of case resin frameworks.

are available on the panel to plug into the unit after it is in place on the slides.

VACUUM TUBES

The tubes have been fully described in an earlier paper;³ their characteristics are summarized in Table I. They are plugged into conventional sockets, and single rather than parallel tubes are used in each stage. These are departures from the practice of the L1 coaxial system, in which two tubes in parallel were soldered in each stage. The use of sockets increases the parasitic capacities and reduces the obtainable feedback by one or two db, but it was felt that the resulting maintenance economy was worth this sacrifice. With single tubes, the failure of one tube results in a line failure and a switch to the protection line. At first sight, it would appear that using parallel tubes in each stage should greatly decrease the probability of line failure. A study of L1 experience, however, showed that most tube failures could either be forestalled by preventive maintenance, or else were of such a nature (for example, shorts within the tube) that the parallel tube would not afford protection against line failure. The reliability advantage of parallel tubes, then, turns out to be small; their use, on the other hand, increases the wiring

TABLE I — TUBE CHARACTERISTICS, DESIGN CENTER VALUES,
NEW TUBES

	435A Tetraode	436A Tetrode	437A Triode
Heater voltage.....	6.3	6.3	6.3 volts
Heater current.....	0.3	0.45	0.45 amperes
Plate-cathode voltage.....	180	180	150 volts
Screen-cathode voltage.....	150	150	— volts
Grid-ground voltage.....	9.0	9.0	9.0 volts
Cathode bias resistor.....	630	315	262 ohms
Plate current.....	13.1	23.4	40.2 milliamp
Screen current.....	3.2	8.6	— milliamp
Transconductance.....	16.5	32.0	47.0ma/volt
Plate resistance.....	—	—	970 ohms
Grid-cathode bias.....	1.3	1.1	1.5 volts

Capacitances (Approximate hot values in mmf)

Grid to plate.....	0.02	0.04	3.6
Grid to cathode and screen.....	9.2	18.2	16.4
Grid to heater.....	0.6	0.8	0.1
Plate to cathode and screen.....	1.0	1.5	0.7
Heater to cathode and screen.....	5.4	7.2	5.4
Heater to plate.....	1.7	1.9	0.2

*Modulation**

Second order (2F).....	36	—	36 db
Third order (3F).....	66	—	67 db
Third order "effective" (3F).....	60	—	60.5 db
Equivalent noise resistance, ohms.....	460	180	—

* Ratio of product to fundamental at output for a 0.1 volt rms signal from grid to cathode.

complexity greatly when sockets are used. The small gain in reliability would not compensate for the degradation in performance caused by the complication of wiring and the resulting capacity penalties.

The applied voltages and current drains are shown in Table II.

HEAT DISSIPATION

In common with many modern designs with high gain vacuum tubes, the problem of heat dissipation was acute in the L3 amplifier. The amplifier is enclosed in a sealed housing of sufficient size to dissipate readily the heat generated. However, with the usual types of chassis and tube shield construction, vacuum tube envelope temperatures were so high that long tube life could not be assured. Consequently a new type of tube shield was developed. This shield is of heavy copper tubing equipped

TABLE II — POWER REQUIREMENTS OF AMPLIFIER UNIT

6.3 volts (grounded).....	1.5 amperes
6.3 volts (190v off ground).....	0.45 amperes
+100 volts regulated	0.5 milliamps
+190 volts.....	68 milliamps
+315 volts.....	41 milliamps

with internal helical springs mounted in such a manner that each turn of the spring makes contact with both the glass tube envelope and with the copper tube as shown in Fig. 6. Thus each turn of the spring provides two metallic heat conducting paths to carry away the heat from the tube envelope. A total of 480 conducting paths are provided for each of the large tubes used in the output amplifier by the use of these springs. Good contact is made between the copper tubes and the chassis, so that heat generated within the vacuum tube is efficiently conducted through metallic paths to the copper chassis of the amplifier. In order not to raise the temperature of the chassis and consequently the temperature of the circuit components, large ribs are provided in the housing, with which the chassis make intimate contact. A continuous metallic con-

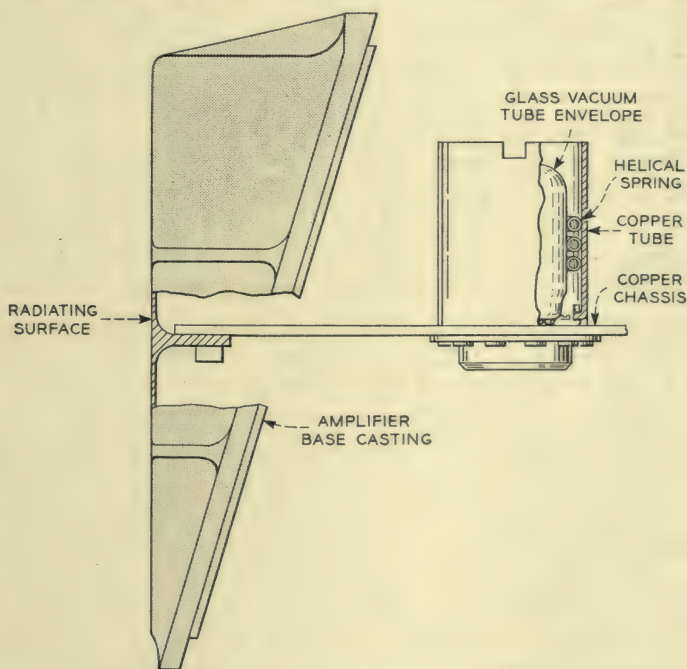


Fig. 6 — Heat conducting tube shield.

ducting path is thus provided from vacuum tube envelope to amplifier housing. This design resulted in a temperature drop in the output amplifier tubes of about 70°C without any temperature rise in the chassis or circuit elements over that produced when ordinary types of shields were used. The smaller vacuum tubes in the input amplifier do not ordinarily run as hot as those in the output amplifier and a temperature drop of only about 55°C was attained by these methods.

COUPLING NETWORKS

The input and output coupling networks are essentially identical. The low side of each transformer is a balanced center-tapped winding which together with the balancing network acts as a hybrid, to produce a good 75-ohm impedance facing the cable. The use of this type of connection gives a signal-to-noise advantage over the use of a brute-force high-side terminating network. The advantage is theoretically 3 db in the case of the output coupling network and would approach the same figure in the case of the input network if tube noise were dominant over the resistance noise of the cable.⁴ Aside from the fact that the design of a high-side shunt termination network for an off-ground peaked transformer is well-nigh impossible, the use of a balancing network in a hybrid connection has the important additional advantage that the adjustment of this network to obtain a good reflection coefficient has negligible effect on the insertion gain of the circuit.

A relatively modest share of the total shaping required has been allocated to the coupling networks: 5.5 db each, or 11 db total. One reason for this is that although these networks are outside the feedback loops in the usual sense, nevertheless the impedances which they present to the amplifiers are important factors in the feedback design, and the effects which they produce must not be allowed to become so severe as to limit the feedback to too low a value. It is obvious from inspection of Fig. 1 that only a part of the voltage developed across the input beta circuit by the plate current of the second tube will appear as a grid-cathode voltage to drive the first stage. The proportion of the beta circuit voltage which will be thus effective in producing feedback around the loop will be dependent on the potentiometer division between the impedance of the coupling network and the grid-cathode impedance of the first tube. The greater the peaking of the input coupling network, the greater its impedance at high frequencies where the grid-cathode capacitive impedance is already decreasing, and hence the greater the potentiometer term loss. A similar loss occurs in the output amplifier. The plate current of the output stage divides between the output

coupling network and the parasitic admittances to ground. The portion of the plate current which returns directly to cathode (ground) through the latter path, without passing through the beta circuit, is not effective in producing loop feedback.

A second and even more important limitation is that the sensitivity of the insertion gain to variations in the coupling network elements is increased as the slope is increased. The same considerations lead us to keep not only the slope but also the gain level or efficiency of these networks relatively low. The maximum possible coupling network gain which can be obtained over the entire frequency spectrum is limited by the capacity across the circuit, as shown by Bode's Resistance Integral Theorem. This capacity cannot be reduced without incurring a more severe potentiometer term penalty and thus limiting feedback, so that the total gain area cannot be profitably increased. The in-band gain can be made greater or less as we concentrate more or less of the total gain area in the transmitted band, but it is found that attempting to get high values of in-band gain area leads to networks which are increasingly sensitive to element deviations. A resistance area efficiency of about 50 per cent turns out to be the most acceptable compromise.

In spite of these steps, the coupling networks are the most important source of manufacturing deviations, but by improved mechanical design and the use of quality control these effects have been reduced by an order of magnitude as compared to previous designs. For example, the end-capacity of each coupling network is a quartz-disc condenser, and the peaking or splitting coils are tension-wound on ceramic forms to give inductances of 1 per cent tolerance with distribution requirements within this range. The windings of the transformers, shown in Fig. 7, are made by plating the turns on threaded forms machined from optical grade quartz or Vycor glass to tolerances of tenths of a thousandth of an inch. Leakage inductance and parasitic capacity deviations are thus held to a minimum. Split ferrite cores, which must also be held to close tolerances, make it possible to use these methods of fabrication, which previously available tape cores would not have permitted.

Since the amplifier configuration gives series feedback on the tubes adjacent to the cable, and cathode feedback on the tubes adjacent to the regulating network, the high impedance winding of each coupling network is necessarily off-ground. This leads to considerable difficulty in specifying the equivalent circuit. For an on-ground coupling network, the equivalent circuit of Fig. 8(b) would be adequate for gain and feedback computations — essentially a reactive equalizer plus an ideal transformer. Even then the capacities and inductances associated with the

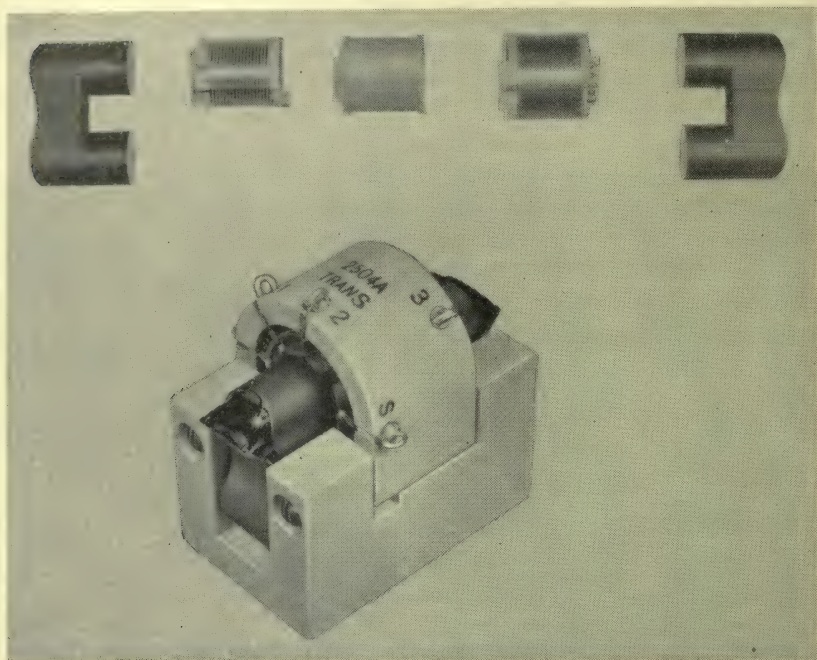
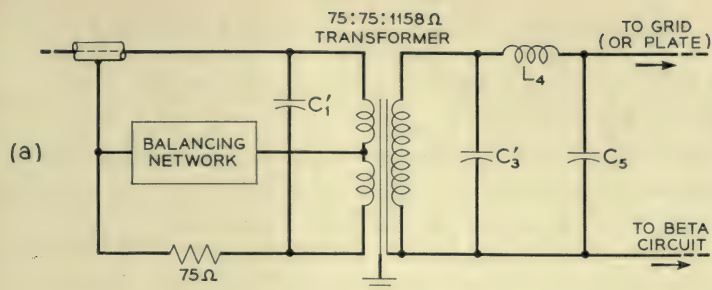


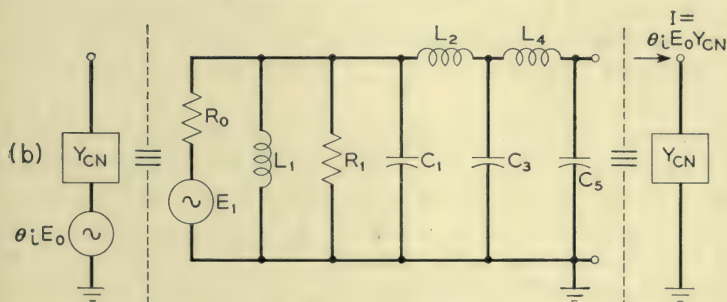
Fig. 7 — Precision transformer, windings, core, assembled unit.

transformer itself would have to be given as functions of frequency because of the distributed nature of the device. When, however, the high impedance winding is raised above ground potential by the voltage developed across the beta circuit, which is nearly the same magnitude as the voltage across the coupling network, the effects of distributed parasitic capacities to ground, and the lumped capacity from the junction of transformer and peaking coil, become of prime importance. The coupling network, therefore, cannot be adequately represented by merely lifting the circuit of Fig. 8(b) off-ground.

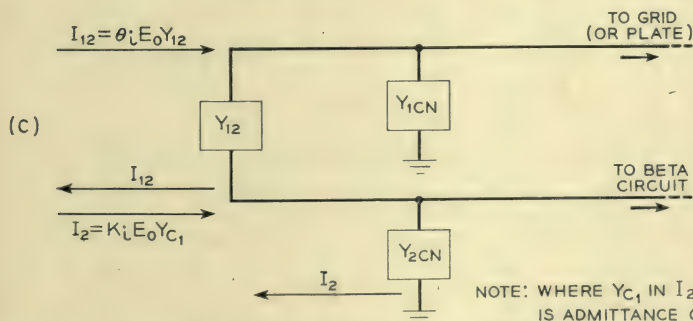
In order correctly to understand and compute the amplifier gain, it was necessary to develop a complex mathematical analysis of the distributed structure of the transformer, in conjunction with an extended program of precise measurements of the transformer constants. Even then, one must be content with an accuracy of a few tenths of a db and a few degrees of phase, as compared with the order of magnitude better accuracy which can be obtained for a two-terminal lumped-constant network. The agreement between measurement and computation of amplifier gain and reflection coefficient is sufficiently good, however, to



- C_1' LOW SIDE PADDING CAPACITANCE
 C_3' HIGH SIDE PADDING CAPACITANCE
 L_4 PEAKING COIL
 C_5 PEAKING CAPACITANCE



- E_1 EQUIVALENT GENERATOR, $\sqrt{\frac{1158}{150}} \times$ CABLE OPEN CIRCUIT VOLTAGE, E_0 .
 R_0 1158Ω.
 L_1, R_1 MUTUAL INDUCTANCE, DISSIPATION OF TRANSFORMER.
 C_1 LOW SIDE CAPACITY $\times \sqrt{\frac{150}{1158}}$
 L_2 LEAKAGE, REFERRED TO HIGH SIDE OF TRANSFORMER.
 C_3 HIGH SIDE CAPACITY OF TRANSFORMER (INCLUDING PADDING CONDENSER).
 L_4, C_5 PEAKING ELEMENTS.



NOTE: WHERE Y_{c1} IN I_2 FORMULA IS ADMITTANCE OF CABLE

Fig. 8 — Coupling Network Circuits. (a) Physical elements. (b) On-ground equivalent circuit, adequate for gain and feedback computations in an amplifier configuration employing on-ground coupling networks. (c) Off-ground equivalent circuit.

assure that all the important effects are sufficiently understood to make intelligent control possible.

Using the results of this analysis, the off-ground coupling network can be represented by the equivalent circuit of Fig. 8(c), where the values of the π of admittances are obtained in terms of the fundamental parameters of the transformer and associated elements. This representation is convenient for insertion gain and feedback computations. The most important difference between the equivalent circuits of Figs. 8(b) and 8(c), from a practical standpoint, is the presence in the latter of the admittance Y_{1CN} which is a manifestation of the capacity to ground from the high impedance winding and the junction of transformer and peaking coil. I_2 and the contributions to Y_{2CN} not directly attributable to the obvious shield-to-shield capacity of the transformer can be set equal to zero with little error, but Y_{1CN} affects the insertion gain and feedback by about 2 db.

It will be observed that the transformer is shown with two shields, one connected to the bottom of the high impedance winding, which is the top of the beta circuit, the other connected to ground. The first of these, which is physically adjacent to the high impedance winding, acts to collect, and carry to the beta circuit, the distributed capacity of the high winding, thus avoiding intolerably large capacity from this winding to ground. The second shield prevents capacitive coupling of the large beta circuit signal voltage to the low impedance winding, which would lead to a very poor reflection coefficient performance. Typical curves of reflection coefficient are shown in Fig. 9. Since the amplifier is an active device, the reflection coefficient is to some extent a function of the vacuum tube transconductances, and tends to be degraded as tubes age.

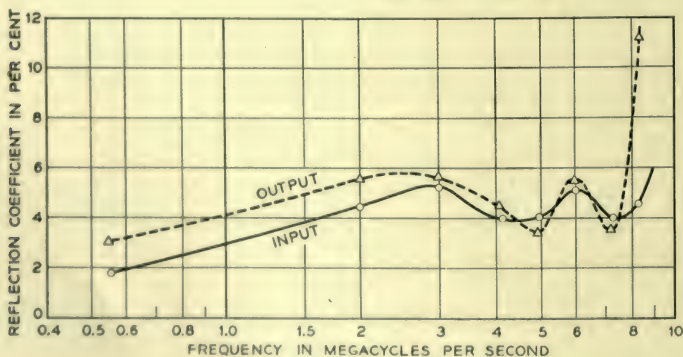
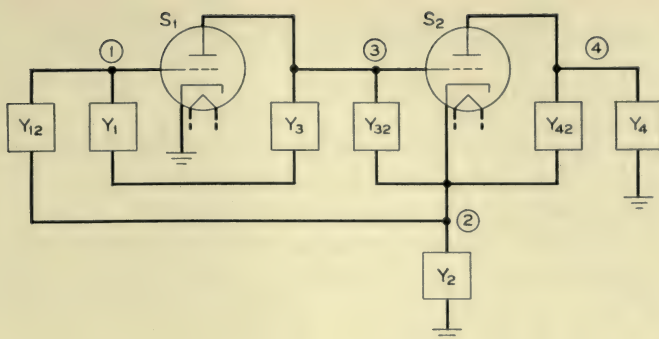


Fig. 9 — Reflection Coefficients.



NODAL DETERMINANT OF CIRCUIT:

$$\Delta = \begin{vmatrix} E_1 & E_2 & E_3 & E_4 \\ Y_{11} & -Y_{12} & 0 & 0 \\ -Y_{12} & Y_{22} + S_2 & -Y_{32} - S_2 & -Y_{42} \\ S_1 & -Y_{32} & Y_{33} & 0 \\ 0 & -Y_{42} - S_2 & S_2 & Y_4 + Y_{42} \end{vmatrix} \begin{matrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{matrix}$$

WHERE $Y_{11} = Y_1 + Y_{12}$; $Y_{22} = Y_2 + Y_{12} + Y_{32} + Y_{42}$; $Y_{33} = Y_3 + Y_{32}$

FEEDBACKS

$$\text{RETURN RATIO ON FIRST TUBE: } T_1 = \frac{P_{12} \frac{S_1 S_2}{Y_{33} Y_2'} P_4 + P_{12} \frac{S_1}{Y_2'} P_{32}}{1 + P_3 \frac{S_2}{Y_2'} P_4} = \frac{B+D}{1+E}$$

$$\text{ON SECOND TUBE: } T_2 = \frac{P_{12} \frac{S_1 S_2}{Y_{33} Y_2'} P_4 + P_3 \frac{S_2}{Y_2'} P_4}{1 + P_{12} \frac{S_1}{Y_2'} P_{32}} = \frac{B+E}{1+D}$$

$$\text{WHERE } P_{12} = \frac{Y_{12}}{Y_{11}}; P_4 = \frac{Y_4}{Y_4 + Y_{42}}; P_3 = \frac{Y_3}{Y_{33}}; P_{32} = \frac{Y_{32}}{Y_{33}}$$

$$Y_2' = Y_2 + \sigma_1 + \sigma_3 + \sigma_4; \sigma_1 = \frac{Y_{12} Y_1}{Y_{11}}; \sigma_3 = \frac{Y_3 Y_{32}}{Y_{33}}; \sigma_4 = \frac{Y_4 Y_{42}}{Y_4 + Y_{42}}$$

TRANSMISSION

$$E_4 = I_{12} \frac{\Delta_{14} - \Delta_{24}}{\Delta} + I_2 \frac{\Delta_{24}}{\Delta}$$

USING FOR I_{12} AND I_2 THE VALUES GIVEN ON FIGURE 8(C)

$$E_4 = E_{oc} \left[\theta_i \frac{Y_2}{Y_4} \frac{B-D \frac{\sigma_4}{Y_2} - E \frac{\sigma_1}{Y_2} - \frac{\sigma_1}{Y_2} \frac{\sigma_4}{Y_2'}}{1+B+D+E} + K_i Y_{c1} \frac{1}{Y_4} \frac{B+E + \frac{\sigma_4}{Y_2'}}{1+B+D+E} \right]$$

Fig. 10 (a) — Input amplifier formulas.

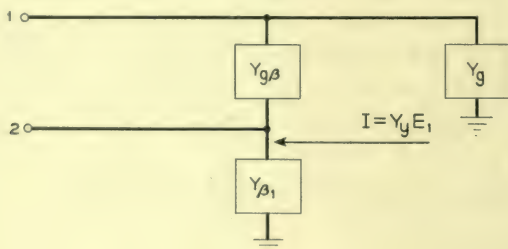
AMPLIFIER FORMULAS

Using the equivalent circuit of the coupling network shown in Fig. 8(c), the circuit of the input amplifier can be represented as on Fig. 10a, and the formulas shown can be derived from straight forward nodal analysis of the circuit. Similar formulas can be derived for the more complicated output amplifier. From these, using Thevenin's theorem, the gain of the tandem combination can be computed, as well as the feedbacks on the various tubes.

Similarly the input amplifier can be replaced, for convenience in reflection coefficient calculations, by the pi of admittances and the driving current of Fig. 10(b). The formulas of Fig. 10(b) expressed in terms of the co-factors of the circuit determinant are of general application for the reduction of a multi-node circuit to simpler form.

REGULATING NETWORK

Like the coupling networks, the regulating network between the amplifiers is outside the feedback loops, and the gain of the amplifier is very nearly a direct function of the impedance seen looking into the network. This impedance is controlled by a single variable resistance — the thermistor — which is directly heated by the dc output current of the regulator. The output of the regulator, in turn, is a function of the



$$Y_g = \frac{\Delta_{22} - \Delta_{21}}{\Delta_{12} - \Delta_{12}} = Y_1 \quad Y_{\beta 1} = \frac{\Delta_{11} - \Delta_{21}}{\Delta_{12} - \Delta_{12}} = Y_2 + \sigma_3 + \sigma_4 + P_3 S_2 P_4$$

$$Y_{g\beta} = \frac{\Delta_{21}}{\Delta_{12} - \Delta_{12}} = Y_{12} \quad Y_g = \frac{\Delta_{12} - \Delta_{21}}{\Delta_{12} - \Delta_{12}} = - \left[P_{32} S_1 + \frac{S_1 S_2}{Y_{33}} P_4 \right]$$

WHERE THE CO-FACTOR Δ_{12-12} IS FOUND BY STRIKING OUT THE FIRST AND SECOND ROW AND FIRST AND SECOND COLUMN OF THE CIRCUIT DETERMINANT, THE SIGN FOLLOWING THE USUAL RULES FOR THE SIGN OF A CO-FACTOR

Fig. 10 (b) — Equivalent circuit of input amplifier as seen by coupling network.

magnitude of main pilot level at the output of the amplifier. The whole forms an AVC circuit which acts to keep the level of the main pilot constant at amplifier output, under the assumption that any changes in this pilot level are caused by line temperature or length deviations. The invariant arms of the regulating network are designed to give the accurate shaping of network impedance versus frequency required so that the amplifier gain change will have the wanted square-root of frequency shape. The design of the network is fundamentally based on the variable equalizer theory developed by H. W. Bode,⁵ but the process of finding physical networks which will give the desired performance to a very high order of accuracy makes use of newer methods developed by S.

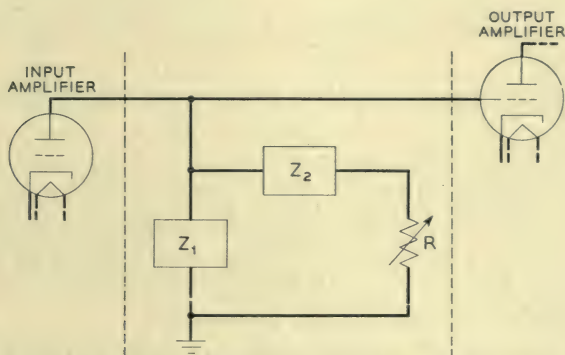


Fig. 11(a) — Regulating network block schematic.

Darlington.^{6, 7} Since these methods were also used in the design of several other amplifier networks, and because the precision thus made possible is essential to the system, a brief recapitulation of the steps involved is of interest.

Fig. 11(a) shows the basic configuration of the regulating network. It was selected because it is one of the simplest that gives symmetrical gain changes controlled by a single resistor. It is also capable of absorbing the parasitic interamplifier capacity, and gives some advantageous gain shaping at normal setting, although this shaping is not under design control.

We need to find the impedance of the two arms Z_1 and Z_2 , that will give the required square-root of frequency shape of gain change (with a value of 6 db at 8 mc) as the thermistor is varied by a factor of three. For this circuit the change in gain may be expressed as a ratio of two impedances.

$$e^{\alpha+jb} = \frac{Z_H}{Z_N} = \frac{\frac{1}{Z_1} + \frac{1}{Z_2 + R}}{\frac{1}{Z_1} + \frac{1}{Z_2 + kR}} = F(p) \quad (1)$$

$$e^{-\alpha-jb} = \frac{Z_L}{Z_N} = \frac{\frac{1}{Z_1} + \frac{1}{Z_2 + R}}{\frac{1}{Z_1} + \frac{1}{Z_2 + \frac{R}{k}}} = \frac{1}{F(p)} \quad (2)$$

where:

Z_N is the impedance looking into the network with a thermistor value R

Z_H is the impedance looking into the network with a thermistor value kR

Z_L is the impedance looking into the network with a thermistor value R/k

For this design $R = 500$ ohms and $k = 3$

We may solve equations (1) and (2) and obtain expressions for Z_1 and Z_2 . However, for synthesis Z_1 and Z_2 should be expressed in terms of singularities in the p -plane. This may be accomplished by approximating the wanted gain change with a Tchebycheff series as developed by S. Darlington. The steps in the process are:

1. The coefficients of a Tchebycheff series that matches the desired gain change over the frequency band of interest are found:

$$\alpha = C_0 + C_2 \cos 2\phi + C_4 \cos 4\phi \cdots C_{2n} \cos 2n\phi \quad (3)$$

where the relation between ω and ϕ is given by the band-pass transformation

$$\omega^2 = K \frac{k_1 + \sin^2 \phi}{k_2 - \sin^2 \phi}$$

$$K = \omega_{c1}\omega_{c2}$$

$$\omega_{c1} = K \frac{k_1}{k_2} \text{ lower cutoff frequency (.2 mc)}$$

$$\omega_{c2} = K \frac{k_1 + 1}{k_2 - 1} \text{ upper cutoff frequency (8.35 mc)}$$

and

$$C_{2k} = \frac{2}{n+1} \sum_{r=1}^{r=n+1} \alpha_r \cos 2k\phi r \quad \begin{array}{l} k = 1, 2, \dots, n \\ r = 1, 2, \dots, n+1 \end{array}$$

(Match points)

$$\phi_r = \frac{(2r - 1) \pi}{(2r + 1) 2}, \quad \alpha_r = \text{gain at } \phi_r$$

2. The coefficients of a related power series are determined:

$$e^{2\alpha} = 1 + A_2 Z^2 + A_4 Z^4 + \cdots A_{2n} Z^{2n} \equiv F(Z^2) \quad (4)$$

where

$$A_2 + S_2 = 0 \text{ and } S_{2k} = kC_{2k}$$

$$2A_4 + A_2 S_2 + S_4 = 0$$

$$nA_{2n} + A_{2n-2} + \cdots = 0$$

3. $F(Z^2)$ is expressed as a rational fraction containing both natural modes and infinite loss points.

$$F(Z^2) = \frac{N(Z)^2}{D(Z)^2} \quad (5)$$

where the coefficients of N and D may be found from the continued fraction expansion of $F(Z^2)$. The degree of N and D fixed by the allowable approximation error and the complexity of the network.

4. The roots of N and D (in terms of Z^2) are found and then transformed to the p -plane using the transformation

$$p_\sigma^2 = K - \frac{2K(k_1 + k_2)}{\psi + 2k_2 - 1} \quad \text{where} \quad \psi = \frac{1 + Z^4}{2Z^2} \quad (6)$$

These four steps result in a polynomial $F(p)$ satisfying the requirements on the change in gain. In this specific case

$$F(p) =$$

$$\frac{1.06(p + 0.0960)(p + 0.0280)(p + 0.9890)(p + 0.2890)}{(p + 0.1058)(p + 1.803)(p + 0.0300)(p + 0.3492)} \doteq e^{\alpha + j\beta} \quad (7)$$

Solving (1) and (2) obtain

$$Z_2 = \frac{k - F}{kF - 1} R, \quad Z_1 = \frac{(k^2 - 1)(F^2 - 1)}{(kF - 1)(k - F)} \quad (8)$$

where $F(p) \equiv F$

Since the design must absorb the interstage capacities, (and also the stray capacities of some of the elements in the physical network) Z_1 must include a shunt condenser. It can be shown that the desired result is obtained when

$$F \rightarrow 1 + g \frac{1}{p} \text{ as } p \rightarrow \infty \quad (9)$$

Since the derived transmission function F , equation (7), does not satisfy the condition expressed by (9), the function must be modified in order to provide a shunt condenser in Z_1 . To accomplish this we use equation (5) expressed as a bi-linear form of the n th term of the continued fraction expansion.

$$F(Z^2) = \frac{N_1(Z^2) + K_n N_2(Z^2)}{D_1(Z^2) + K_n D_2(N^2)} \tag{10}$$

The coefficient K_n is modified so that the new $F(p)$ satisfies equation (9). The modification of K_n does not substantially alter the required in-band transmission and also does not produce non-physical singularities. The new F obtained is

$$F(p) = \frac{(p + 0.0289)(p + 0.1018)(p + 0.3431)(p + 1.017 \pm i\,1.628)}{(p + 0.0310)(p + 0.1132)(p + 0.4426)(p + 0.6158 \pm i\,1.368)} \tag{11}$$

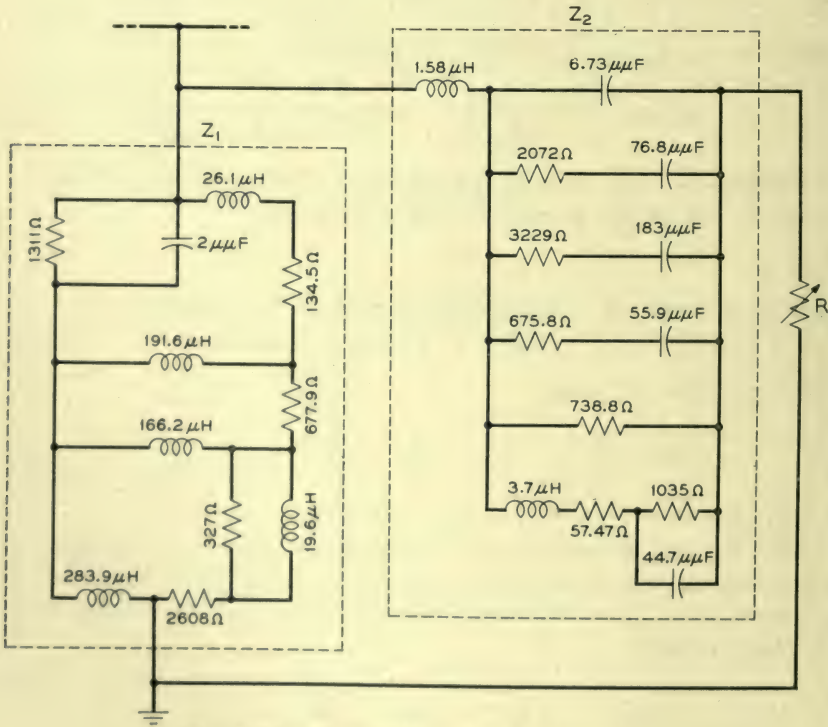


Fig. 11(b) — Regulating network schematic.

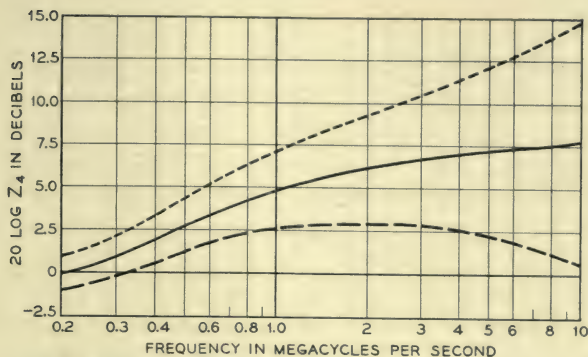


Fig. 12 — Relative gain of regulating network for extreme and mid-range thermistor settings.

By using this result in equation (8) expressions for Z_1 and Z_2 were obtained from which the configuration that is shown in Fig. 11(b) was synthesized.

Fig. 12 shows the voltage which would be developed across the regulating network versus frequency in response to a constant current driving force, for the mid-range and extreme values of thermistor resistance. The slope across the band for the mid-range value of 7.5 db; this is the regulating network contribution to the total slope of 37 db required of the complete amplifier.

To the extent that the differences between the curves of Fig. 12 fail to exactly match the desired square-root of frequency characteristic, the action of the regulating network will introduce an equalization error. This regulation error is shown in Fig. 13. It amounts to a few hundredths of a db for a six db gain change, caused in part by network design imperfections and in part by the fact that very small second order effects result in the amplifier gain not exactly following the regulating network impedance.

BETA CIRCUITS

Starting from our basic requirement that a slope of 37 db across the transmitted band must be obtained, and noting that the total of the contributions from the coupling networks and regulating network is 16 db, we are left with about 20 db of shaping to be supplied by the beta circuits. The input beta circuit has been designed to supply most of this remainder, the contributions of the output beta circuit and the $\mu\beta/(1 - \mu\beta)$ effects in the amplifiers being only three db. The original design procedure for this network was basically the same as for the

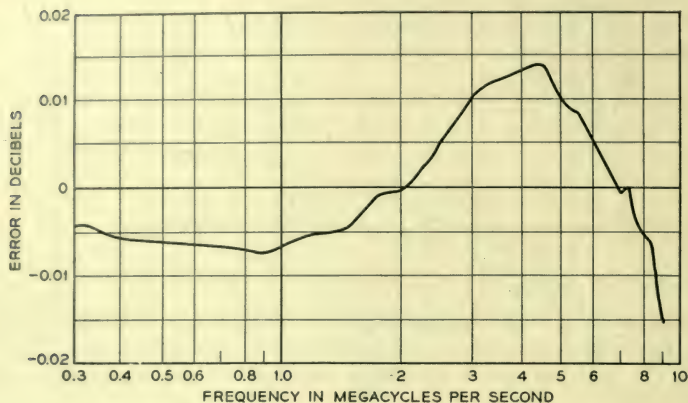


Fig. 13 — Regulation error per one db regulation.

regulating network, but the final values of the elements represent a long process of incorporating secondary corrections as our knowledge of the amplifier grew. Original constraints included the necessity of including as part of the beta circuit not only the relatively large parasitic capacity of the amplifier and the coupling network shield to shield capacity, but also the screen resistor and by-pass condenser of the second tube, which as usual is by-passed not to ground but to cathode, for modulation reasons. It is also required that the beta circuit have the correct dc resistance to serve as the cathode bias resistor of the second tube, and that it incorporate provision for metering this bias through suitable decoupling elements which are among the gain-determining elements of the network. Finally, this network was used as the mop-up equalization network of the amplifier, and its element values were readjusted to give the correct amplifier gain after the performance of a representative group of amplifiers with average coupling networks and tubes had been determined. In doing this tailoring, it is necessary to precorrect for the effects of non-infinite feedback, since the gain of the amplifier is not exactly a direct function of the beta circuit admittance. The configuration of the input beta circuit is shown by Fig. 14; it is a simple two terminal network. Its in-band impedance varies from about 1,000 ohms at 300 kc to 110 ohms at 8.35 mc.

The output beta circuit is relatively flat, which in this case is the optimum condition for signal-to-noise and feedback loop stability considerations. Because of this simplicity, it is possible to incorporate in this network provision for adjusting the gain of the amplifier to reduce the misalignment of the system.

MISALIGNMENT ADJUSTMENT

We can distinguish three major effects which will contribute to misalignment and consequent degradation of system signal-to-noise ratio. One is design error — the degree to which the design gain of the amplifier, because of the finite number of elements and other limitations, fails to match the line loss. Second is the cumulative effect in the individual amplifier of manufacturing deviations of the elements. Third is the aging of the components of the amplifier, of which the tube aging will of course be the dominant short term effect. The signal-to-noise performance of the system can be improved by reducing the misalignment, if this is done without resorting to measures which would introduce systematic instead of random gain deviations.

If we study the shapes of gain change introduced by the more important deviation contributors, particularly the element variations, we find as might be expected that in general the effects are small at low frequencies and increase sharply near the upper edge of the band if the thermistor is held constant. In the system, the regulation around the main pilot will automatically act to reduce the deviation at 7 mc to zero, adding a square root of frequency curve that results in a bow shaped deviation as illustrated by Fig. 15. Examining the signal-to-noise effects of the degradation caused by misalignment, we conclude that it is most desirable to reduce as much as possible the misalignment at 4 mc, the television carrier frequency in the combined system. The output beta circuit has, therefore, been designed to give varying amounts of this shape, the total range being ± 0.6 db at 4 mc in 0.2 db steps. The successive steps of this gain adjustment are simple multiples of each other, symmetrical in the two directions of adjustment, so that we put into

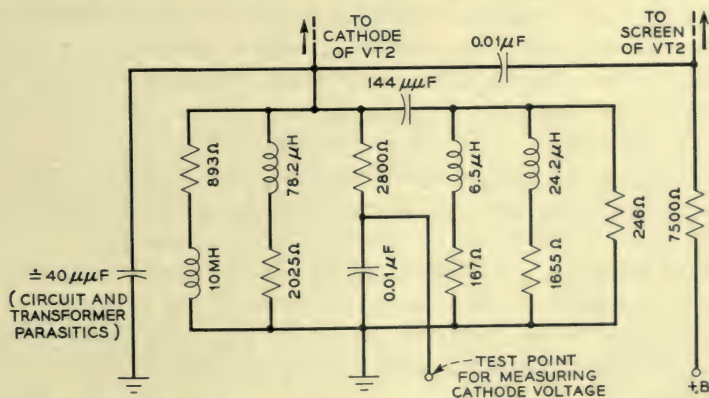


Fig. 14 — Input beta circuit elements — line amplifier.

the system a single systematic shape, which will be small in magnitude at the equalizing points because the different settings of this control in successive line repeaters will tend to cancel out. This setting of output beta circuit gain is controlled by a GAIN ADJ switch accessible from outside the amplifier housing. The adjustment will be made on an in-service basis as part of normal maintenance procedures, using the level at each repeater of one of the system pilot frequencies (3,096 kc) as the index of proper setting.

MANUFACTURING TESTING

As mentioned above, the mechanical design of the amplifier has been planned to permit the separate testing of the five gain-determining net-

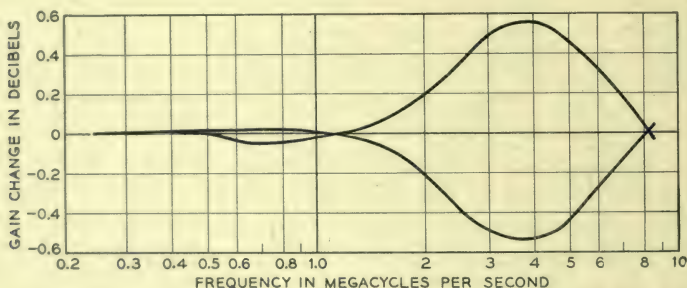


Fig. 15 — Amplifier gain versus output beta circuit misalignment adjustment.

works, the tuned interstage of the input amplifier, and the separate input and output amplifiers before their assembly with the regulating network to form complete amplifiers. Variations in environment are minimized by the use of jigs which also make it unnecessary, in general, to solder to the network under test. Visual gain sets which cover the transmitted band and are accurate to two or three hundredths db are used. The component network or component amplifier under test is connected in series with a complementary or equalizing network to obtain a flat transmission characteristic which can be accurately compared to the transmission of standard attenuators. The completely assembled amplifiers are similarly tested. Quality control charts of the resulting measurements on all components are useful in detecting shifts in transmission which might be caused by loss of control in element manufacture or by shifts of element values caused by subsequent damage in handling.

DC FEEDBACK

In addition to the loop feedback at signal frequencies, local dc feedback is used on each stage. The grid of each tube is returned to a +9-volt potential rather than to ground, and about +10 volts is developed across each cathode resistor. The usual stabilizing effects of self-bias are thus obtained in exaggerated degree, each tube having about 20 db of local dc feedback. Care must then be taken to select the cathode by-pass and interstage coupling condensers so that the transition from low-frequency local feedback to in-band loop feedback is accomplished smoothly without the instability which might be caused by a balancing out of these two feedbacks in the transition region.

For maintenance measurements, the 9-volt bias potential and the dc cathode voltage of each tube are brought out to a multi-pin amplifier test jack through appropriate decoupling filters. Thus the bias on each stage can be measured on an in-service basis. Filament voltage dropping resistors which can be switched in or out of circuit are provided on the repeater panel, so that bias can also be observed for a filament supply voltage 10 per cent below normal. The activity or change in plate current with filament voltage thus measured, or the history of the bias at normal filament voltage, will be used to determine when amplifiers should be taken out of service for tube replacement.

The upper triode is, of course, a special case: since it is the plate supply path for the lower triode, its cathode is about 160 volts above ground and its grid is returned to a similarly high potential. About 35 db of dc feedback is obtained on this stage; the same provisions for measurement of bias and activity are made. To avoid excessive filament to cathode voltage, a separate filament winding which floats at the +190-volt supply potential is used for this tube.

LOOP FEEDBACK

The design of the feedback loops of the input amplifier follows conventional practice. The constraints which operate to limit the amount of feedback which can be obtained in the transmitted band are well known.⁸ Broadly speaking, the figure of merit of the vacuum tubes and the circuit capacities determine the asymptotic cutoff, which in any feedback amplifier limits the magnitude of the feedback which can be built up in the band. In multi-loop structures, there are additional limitations. Consider, for example, the formula given on Fig. 10(a) for the feedback on the second tube. If we increase without limit the magnitude of the first tube transconductance, we find that the feedback

on the second tube approaches the value $S_1 P_4 Z_{32}$. This is a limit, common to multi-loop structures, over and above the usual limit imposed by the Nyquist stability criterion. A similar $S P Z$ limit can be derived for the feedback on each stage. The same multi-loop mechanisms which operate to limit the feedback also result in making the feedback on any given stage relatively insensitive to variations in other transconductances or impedances. In a single loop circuit a one db change in beta circuit impedance or first stage transconductance causes a one db change in the feedback on every tube. Here the feedback on, say, the second stage would generally change only one-half db at most frequencies. While this is an advantage in the sense that transconductance decay does not decrease feedback as rapidly as it would in a single-loop amplifier, it is a disadvantage in that it militates against improving stability margins by shaping the out-band impedance of the beta circuit.

The capacity distributions within the input and output amplifiers are such that while the $S P Z$ limitations on feedback are closely approached, the most stringent limitation on the feedback obtained is the Nyquist criterion.

The use of the Nyquist criterion, particularly with respect to defining the margins against singing, is likewise complicated by the multi-loop nature of the circuit. Ignoring some very recent work, the implications of which have not been fully explored at this writing, it can be said of a multi-loop structure that the apparent margins against singing shown by any plot of feedback give no certain information as to how safe from singing the circuit is. The phase, as well as the magnitude, of the feedback on each stage is a function of the magnitude of the other transconductances, and either decay or increase of these transconductances might destroy the phase margin. In these circumstances, it is theoretically necessary to examine for stability every conceivable combination of transconductances. A more practical expedient, of course, is to rely on judgment backed by computation and laboratory experiment on the circuit for a wide but far from infinite number of circuit conditions.

Another difficulty arises from the fact that the feedback on each tube is different, so that gain and phase margins obtained for one return ratio do not imply equal gain and phase margins for the return ratio on some other tube of the same amplifier. It does not follow, however, that it is necessary to investigate separately the margins on each stage versus circuit element variations. The point to be stressed here is that we are using the behaviour of the return ratio merely as an index of our real concern: the position on the p -plane of zeros of the determinant of the circuit, and the determinant is the same for both stages. Rather

than asking how many db of gain margin and how many degrees of phase margin any particular return ratio displays, we ought instead to inquire how quickly the apparent margins disappear as we change transconductances and network impedances. The margins on all the return ratios will vanish simultaneously, regardless of the apparent difference in original margin magnitudes. It is therefore satisfactory to examine the behaviour of whichever return ratio is most easily observed.

The design choices made in arriving at the coupling network also affect the magnitude and shape of the feedback which can be obtained: as mentioned above, the relative magnitude of the impedance seen looking into the coupling network and the impedance from first tube grid to ground determines how much of the voltage developed across the beta circuit will reach the first stage as a driving force.

Study of the modulation products which will arise in system operation, both for the all telephone case and for the combined telephone-television signal, led to the conclusion that optimum shaping of the feedback for the L3 system would be to maximize the feedback at low frequencies in order to suppress intermodulation products falling in that part of the spectrum in the combined telephone-television case. Shaped feedback, falling off at the higher frequencies, is also consistent with obtaining a smooth and simple shape of gain-change as tubes age (known as "mu-beta effect"), which is desirable from the equalization standpoint. With these considerations in mind, the interstage of the input amplifier has been peaked well above the transmitted band — at 11 mc — to partially compensate for the input potentiometer term, and to help in achieving this smooth shape of mu-beta effect. If flat feedback over the band were the objective, it would also be necessary to design the grid-cathode admittance of the second tube so that the parasitic grid-cathode capacity would be absorbed in a flat impedance versus frequency, but in this case the desired shaping of the second tube feedback is attained by taking advantage of the way in which the grid-cathode capacity naturally limits the high-frequency feedback on this stage.

The loop feedback in the output amplifier is similarly shaped, for the same modulation and equalization reasons. The use of the double triode circuit in this amplifier is an unusual feature. This connection of two triodes, sometimes referred to as the "cascode circuit," has appeared in many contexts in recent years, usually to serve some other purpose than here. It serves as a superior output stage in the L3 amplifier largely because the effective transconductance which can be obtained is about 3 db higher than that of a pentode of the same grid-cathode spacing. The effective transconductance, ignoring for the moment the division of out-

put current between the load impedance and the internal plate impedance of the upper triode, is very nearly the transconductance of the lower triode. Since no current is lost to a screen, this is higher than that of a pentode of the same spacing. The output impedance of the upper triode is multiplied by the local feedback consequent on its cathode being off ground by the plate impedance of the lower triode. Since the load presented to the lower triode is a low impedance, approaching the reciprocal of the transconductance of the upper triode, the grid-plate capacity of the lower triode is not enhanced by feedback. The higher value of transconductance means less grid drive for the same output current, and more obtainable feedback, both contributing to superior modulation performance.

The modulation of the upper triode, and the changes in transconductance of this tube, are suppressed by approximately the sum, in db, of the local and loop feedback. In consequence, the modulation and mu-beta effect contributions of this tube to the complete amplifier are negligibly small.

It will be observed that the grid of the upper triode is not connected to ground, but to the top of the beta circuit. In consequence, the grid-plate capacity of the upper triode appears as part of the end-capacity of the coupling network rather than as a parasitic capacity to ground. This results in a gentle potentiometer term in the output amplifier. This connection, however, also has the effect of vitiating to some extent the desirable qualities of the circuit, particularly at the higher frequencies, where the grid-cathode capacity of the upper triode becomes important.

Because of this gentle output potentiometer term, it is not necessary to tune the interstage of the output amplifier, which consists simply of the circuit capacity plus a network which has the characteristics of a 10-mmfd capacity in the transmitted band. Above the band this network shapes the gain and phase characteristics of the feedback loop to obtain the desired stability margins. The incorporation of this network reduces the in-band feedback by about 3 db, a sacrifice which unfortunately is necessary to assure stability when the thermistor in the regulating network is at its minimum value. For this value of thermistor, the phase and magnitude relations of the regulating network impedance and the grid to cathode capacity of VT3 produce a potentiometer term in the feedback loop which appreciably reduces the margins around 30 mc. The stability margins for the mid-range value of thermistor would be satisfactory without this sacrifice of in-band feedback. When the thermistor is at maximum resistance, some degradation of phase margin at 10 mc occurs, again because of the potentiometer term effect mentioned above,

but in this case the remaining margin is sufficient, since the circuit elements are still under good control at this frequency. Because the plate-cathode impedance of VT_2 is very high, the similar potentiometer term at the output of the input amplifier causes only negligible changes in input amplifier stability margins as the thermistor changes.

In the 70-mc region there are two almost equally important feedback loops in the output amplifier — one through the transconductance of the lower triode, the other through the grid-plate capacity of this tube. Balances between these feedback paths are observed in the 70 to 100-mc region in the course of measuring the feedback, sometimes accompanied by 180° shifts in the phase of the loop transmission at frequencies above the balance point, an effect which theoretically depends on just how the two vectors go through the balance point. The occurrence of these balances is accompanied by a few degrees loss of phase margin in the 30-mc cut-off region, which must also be allowed for in setting the 30-mc stability margins, since sufficient control of parasitics to prevent these 70-mc effects is out of the question.

Parasitic resonances between the lead inductances and the capacities of the circuit, which tend to cause instabilities in the very high-frequency region about 200 mc, are damped by small resistors in the leads, and the lead inductances are kept small by careful mechanical design. In this frequency region, neither measurement nor computation of stability margins can be trusted as anything but a rough guide. On the other hand, adding damping resistors in grid leads and other critical points to prevent 200-mc sings causes a phase margin penalty in the 30-mc region, so a nice judgment of how much damping to add is called for. Final values of damping were chosen so that typical amplifiers could not be made to sing by increasing critical lead lengths or by substantial increases in parasitic capacity, thus assuring that manufacturing variations of elements and wiring will not cause high-frequency sings. The return ratio of VT_4 is shown on Fig. 16, the mu-beta effects of both amplifiers on Fig. 17.

SIGNAL LEVELS, MODULATION, AND NOISE

Fig. 18 shows the signal levels within the amplifier in db relative to one volt from grid to cathode of VT_4 , which is a convenient point to use as a reference for system signal-to-noise studies. It will be noted that as a result of using the input beta circuit to give so much of the shaping of amplifier gain, the input amplifier has little gain at low frequencies. In consequence the input tube of the output amplifier and the regulating network are important thermal noise sources at low frequencies. The

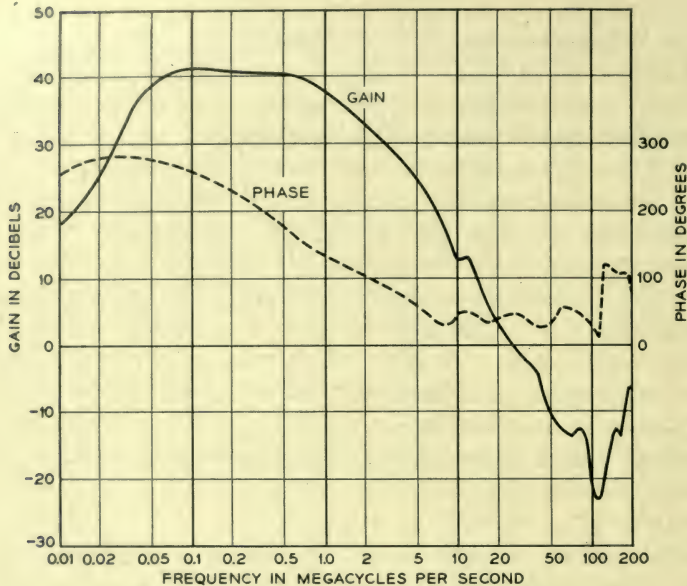


Fig. 16 — T4, return ratio of lower triode.

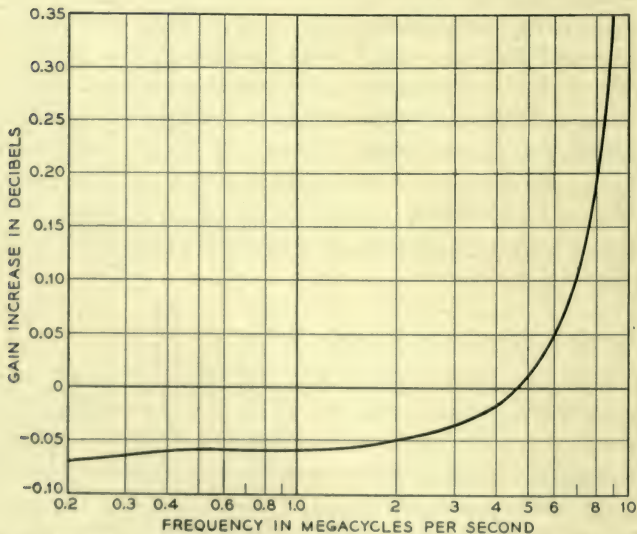


Fig. 17 — Mu beta effect, amplifier gain change for a one db decrease in each transconductance.

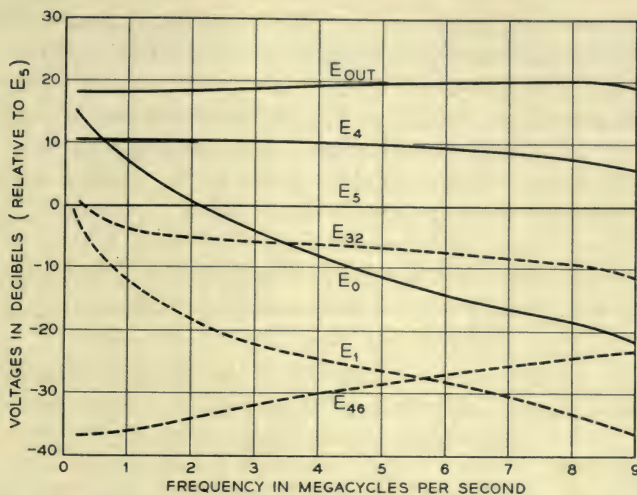


Fig. 18 — Relative levels of signal voltages in line amplifier referred to voltage at grid of lower triode. Thermistor at mid-range value.

relative magnitudes of the noise sources are shown on Fig. 19, which gives the noise at amplifier output as a function of frequency.

Comparison of the grid to cathode voltages of $VT2$ and $VT4$ shows that the former will be an important modulation contributor, since the driving force on these tubes is nearly equal, particularly at low frequencies. Typical amplifier modulation values are given in Table III. Computations using the measured feedback and the performance of

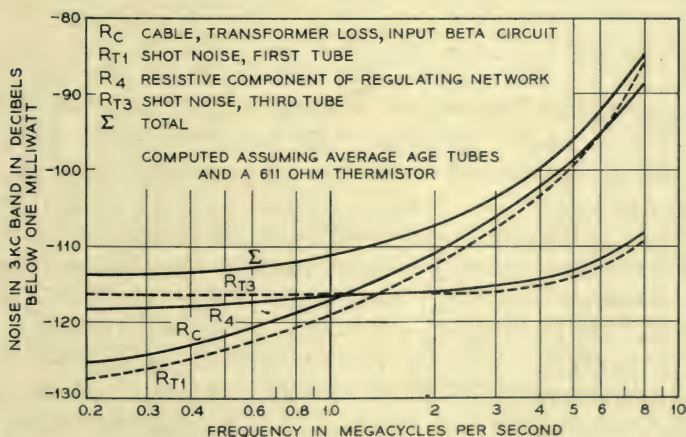


Fig. 19 — Thermal noise at line amplifier output.

single tubes without feedback check the measured values of amplifier modulation to within a couple of db if the third order coefficient of the tube is corrected to take account of the fact that some third order modulation is generated by the interaction of the fundamentals and the feedback second order products. In general, the effective third order coefficient of the tubes is approximately equal to the voltage sum of the tube's uncorrected third order coefficient and a coefficient 6 db worse than the square of the tube's second order coefficient. If this interaction correction is not taken into account, the correlation of tube modulation, feedback and amplifier modulation is unsatisfactory. The analysis leading

TABLE III — MODULATION PRODUCTS, IN db BELOW ONE MILLIWATT AT AMPLIFIER OUTPUT, FOR FUNDAMENTALS 5 db ABOVE ONE MILLIWATT AT AMPLIFIER OUTPUT

Type	Fundamentals Mc	Product Mc	Product -dbm
2F	0.5	1.0	72
	1.0	2.0	65
	3.0	6.0	55
	4.0	8.0	51.5
3F	0.25	0.75	95
	0.667	2.0	97
	2.0	6.0	87.5
	2.66	8.0	81.5

to this result, which is due to F. B. Llewellyn, S. E. Miller and R. W. Ketchledge, is too long to give here.

LOAD CARRYING CAPACITY

The load carrying capacity of an amplifier is difficult to define with exactness. One possible definition is the load at which the modulation coefficients of the amplifier have departed appreciably from the small signal power series values because of loss of feedback as the transconductance is cut off during part of the cycle. The signal carried without serious overload, in terms of a single frequency, is practically constant in the transmitted band as a consequence of the fact that the output voltage and the lower triode grid voltage have nearly the same shape versus frequency, as shown on Fig. 18. The output coupling network shaping approximately compensates for the potentiometer term division of current between the load impedance and parasitic paths to ground. Departure from the small signal power series behaviour just begins to

be appreciable at +14 dbm of any single frequency. At +18 dbm the dc effects of overload show up as slight changes in the transmission of the pilot frequencies. At +26 dbm, the second order modulation is 3 db, the third order modulation is 6 db, higher per line amplifier than would be predicted from small signal behaviour.

FLAT AMPLIFIER

The flat gain amplifier, which is used as a transmitting amplifier and to make up for equalizer loss at various points in the system, is basically the same as the line amplifier, with only the obviously necessary modifications. The input beta circuit is nearly flat, the regulating network has been replaced by a fixed gain network which contains a single variable element whose adjustment at the factory compensates to some extent for variations in the coupling networks, and the input coupling network has been modified so that the peaking used in the line amplifier is replaced by a drop in the high-frequency gain of this network. The output beta circuit has been modified to give flat gain control of ± 1.0 db in 0.2 db steps. The interstage designs are changed to readjust the feedback so that the modulation suppression and the change in gain as tubes age will be nearly the same as in the line amplifier. Somewhat more feedback is obtained in the output amplifier since no network is needed in the output amplifier interstage to adjust the 30-mc phase for an unfavorable regulating network setting.

The nominal gain of the flat gain amplifier has been set at 34 db and is flat to within ± 0.2 db over the transmitted band. The amplifier circuit capacities are low enough so that the inter-amplifier network could be built to give considerably more gain than this; the limit has been set so that flat gain amplifier noise contributions to the complete system noise will not exceed about 1.0 db at the television carrier.

ACKNOWLEDGEMENTS

As in any corporate development, many members of the Laboratories have made important contributions to the L3 amplifier design. Particular mention should be made of the work of S. E. Miller on fundamental amplifier design, S. Darlington and T. R. Finch on network design, C. W. Thulin on the precision transformer, B. J. Kinsburg on quality control problems, E. Ley on mechanical design, and E. F. O'Neill on the flat gain amplifier and on high-frequency stability problems in both amplifiers.

REFERENCES

1. C. H. Elmendorf, A. J. Grossman, R. D. Ehrbar, R. H. Klie, The L3 Coaxial System — System Design, see pp. 781 to 832 of this issue.
2. R. W. Ketchledge, T. R. Finch, The L3 Coaxial System — Equalization and Regulation, see pp. 833 to 878 of this issue.
3. G. T. Ford and E. J. Walsh, Electron Tubes for a Coaxial System, Bell Sys. Tech. J., Oct. 1951, Part II.
4. H. W. Bode, Coupling Networks, U. S. Patent 2337965, Dec. 28, 1943.
5. H. W. Bode, Variable Equalizer, Bell Sys. Tech. J., April, 1938.
6. Sidney Darlington, The Potential Analogue Method of Network Synthesis, Bell Sys. Tech. J., April, 1951.
7. Sidney Darlington, Network Synthesis Using Tchebycheff Polynomial Series; Bell Sys. Tech. J., July, 1952.
8. Hendrik W. Bode, Network Analysis and Feedback Amplifier Design, D. Van Nostrand, 1945.

The L3 Coaxial System

Television Terminals

By JOHN W. RIEKE and R. S. GRAHAM

(Manuscript received April 17, 1953)

Television terminals are required at circuit ends of the L3 coaxial system; at the transmitting end to condition video signals for carrier transmission and at the receiving end to detect the transmitted signals. Special signal characteristics, e.g., a degree of modulation which exceeds the value commonly referred to as 100 per cent modulation, require departures from standard modulating and detecting processes. The high degree of modulation requires both careful control of transmitted wave form and at the receiver product demodulation with phase synchronous carrier (homodyne detection).

Carrier regeneration requirements result in the choice of one step frequency translation from the video frequency spectrum to the allocated vestigial sideband carrier spectrum. The one step process using a single modulator results in unusual balance requirements for the modulator itself and an unusual circuit configuration.

Transmission quality objectives for the terminals are such as to permit six pairs of television terminals to operate in tandem in a transcontinental circuit. This permits a degree of interconnecting flexibility in operation with other systems, e.g., L1 coaxial or microwave systems. These objectives place severe requirements upon the transmission stability of various filters and other circuits within the terminals. New network techniques both in design and fabrication are brought to bear in the effort to achieve required performance.

The transmitting and receiving terminals are described, illustrating the functional operation and mechanical and electrical arrangements of the equipment.

INTRODUCTION

The main features of the L3 coaxial cable transmission system have been described in a companion paper.¹ This paper describes the television transmitting and receiving terminal equipment of the L3 system. The

transmitting terminal conditions a television signal for carrier transmission over the system simultaneously with a group of 600 telephone messages. The receiving terminal reconverts the carrier signal at each receiving point along the cable route. Primarily, the transmitting terminal is a modulator which translates the composite video picture spectrum of frequencies up to the carrier band of frequencies and the receiving terminal is a detector which retranslates the carrier spectrum back to its original band of frequencies.

Particular characteristics of the transmitted television signal, which are intended to aid in achieving optimum transmission quality, have necessitated the departures from past techniques in modulation and demodulation processes that are described in the following. Described also are the methods employed to achieve transmitted picture quality adequate for tandem operation of as many as six pairs of transmitting and receiving terminal equipments in a 4,000-mile television transmission circuit. Operation with several pairs of terminals in tandem occurs when L3 coaxial systems are interconnected with L1 coaxial systems or microwave radio systems.

L3 television terminal development has been in progress since early in 1948. Two transmitting and two receiving terminals have been built on a preproduction basis and currently are being tested under field conditions as part of the L3 system field trial. Development effort is continuing on the terminals with emphasis on equipment reliability, including means for maintaining and improving transmission quality.

FREQUENCY ALLOCATIONS

The L3 coaxial system was designed to have as broad a transmission band as the economics of repeater spacing together with presently realizable feedback amplifier performance permit.² The band extends from 300 kc to 8.5 mc. In comparison with this band a broadcast television signal occupies the frequency spectrum from zero frequency up to 4.5 mc.

From the foregoing it is evident that the television spectrum will not occupy fully the available system transmission band. It is feasible and attractive to allocate part of the transmission band for television transmission and the remainder for transmission of message channels. Detailed allocations then result from a compromise among transmission performance, cost and the number of message channels made available.

From these considerations vestigial sideband transmission of the television signal rather than double sideband transmission is called for. The smaller the vestigial band of transmitted frequencies is made the

smaller will be the total television band required. However, both the cost of band shaping filters and the difficulty of maintaining satisfactorily low values of vestigial sideband quadrature distortion increase as the vestigial band width is reduced. The compromise of these factors resulted in the choice of a 500-kc vestigial sideband.

Another choice made was to transmit the television signal in the upper part of the L3 band and the message channels in the lower part. This allocation was determined by considering the noise distribution in the transmission band together with the modulation distortion, (harmonic distortion), produced by the repeaters. By transmitting television in the upper part of the band a minimum modulation distortion is achieved since the harmonics of the television signal largely fall outside the transmitted band or at high frequencies where their effects in the picture are relatively less visible than low frequency distortion. This factor outweighs the higher noise level in the upper part of the band.

With respect to the television carrier location, it is placed at the bottom of the television band at 4.139 mc, with the vestigial sideband extending down to 3.64 mc and the main sideband extending upward to 8.50 mc. Alternatively the carrier could have been located at the top of the L3 band with a main lower sideband and vestigial upper sideband, but this choice would be disadvantageous because of higher noise levels and poorer repeater gain stability near the top edge of the transmitted band.

The final allocation of frequencies is shown in Fig. 1. Just below the television band from about 3 mc to 3.5 mc is a dead space. This frequency space is needed for the filters which are employed to separate the telephone signals from the television signals at television and telephone dropping points. A very high value of loss is required of these filters in their attenuating bands and if this is built up over too small a frequency

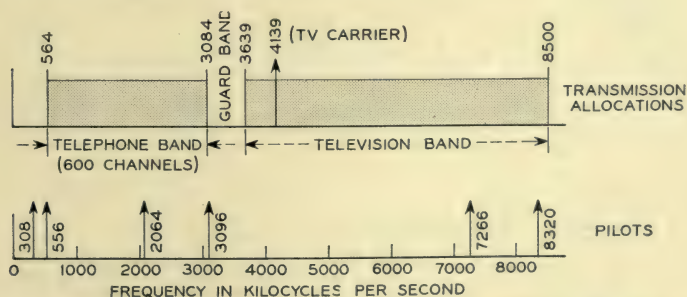


Fig. 1 — Frequency allocations for L3 combined television-telephone transmission.

band the resulting delay distortion introduced into the television band becomes very difficult and expensive to equalize. Below this "guard band" is the "master group" of 600 telephone channels which is transmitted simultaneously with the television signal.

The detailed allocation of specific frequencies, e.g., TV carrier and the pilot frequencies, results from consideration of effects produced by these frequencies in telephone channels as a consequence of modulation distortion in repeaters. These considerations are described in detail in the system design paper.¹

MODULATION PROCESS

In the L1 coaxial cable system³ frequency translation by the television terminals is accomplished in two stages. The two step process employs a first modulator supplied with a very high frequency carrier to translate all video frequencies to a band far outside the final transmitted band of

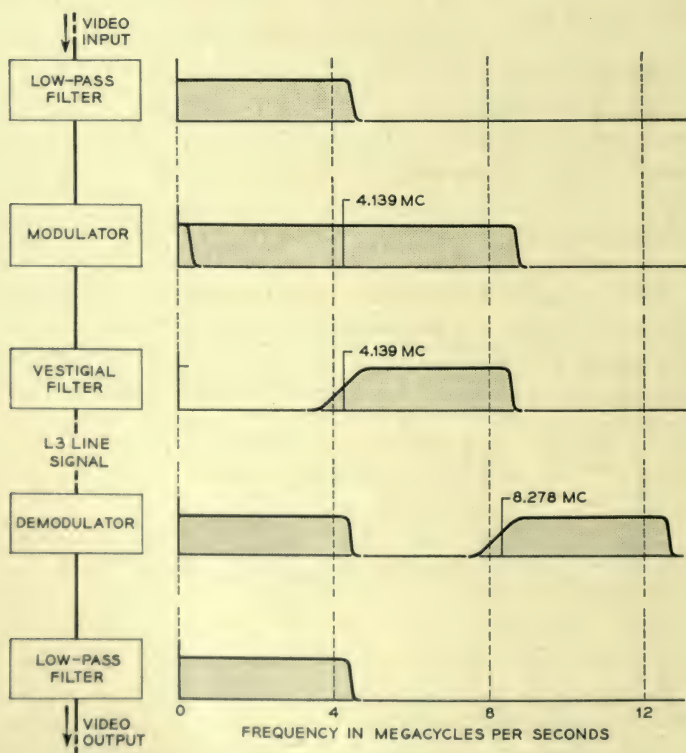


Fig. 2 — Television terminal modulation processes.

frequencies where the upper side band is suppressed. A second modulation then translates the vestigial sideband signal back down in frequency to the final band. In contrast the L3 terminals employ only a single step of modulation to convert the signal directly to the assigned band as shown in Fig. 2. In general this can be accomplished if the carrier frequency is at least half the sum of input and vestigial bandwidths. Then the lower modulation sideband does not fold over the zero frequency axis to produce frequencies which fall back into the vestigial or upper sidebands. Some foldover is evident in the L3 case shown in Fig. 2 at low frequencies of the modulator output. Single step modulation is advantageous in that the very high frequencies encountered in the multi-step process are avoided. The disadvantage of the one step process is that many extraneous products of modulation, which in the two step process can be suppressed with filters, must be reduced to tolerable levels by balances in the modulator.

The modulator, Fig. 3, is a combination of two double balanced modulators of a form often employed for modulation of telephone signals.⁴ The effect of a double balanced varistor modulator of the type represented by either of the two in Fig. 3 is to multiply the input signal by a square wave function having the period of the particular carrier frequency. Since the square wave contains all odd harmonic multiples of

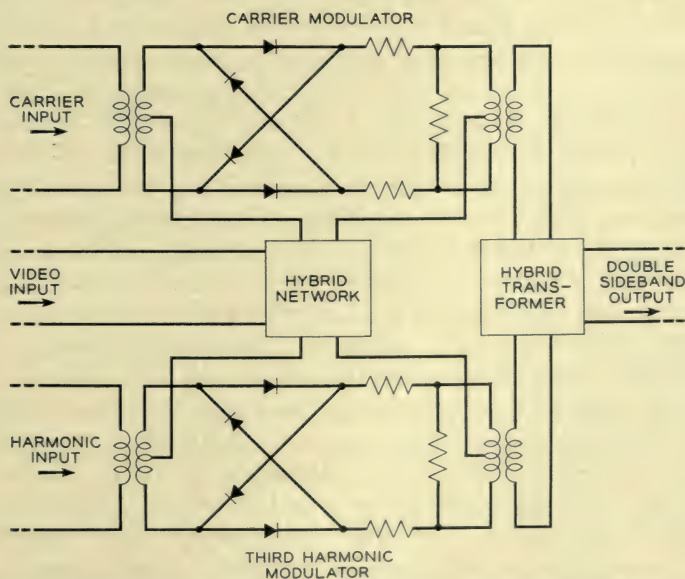


Fig. 3 — Modulator for L3 terminals with third harmonic carrier product balance feature.

the carrier frequency its multiplication by the input signal generates a series of double sideband output spectra, each centered about one of the harmonics of the carrier. It happens in this case that the lower sideband of the third harmonic spectrum of the carrier modulator contains frequencies low enough to overlap the high frequencies of the carrier spectrum upper sideband and this overlap results in quite visible picture distortion.

It is possible, by employing a second modulator paralleling the first but driven with a frequency three times the carrier frequency, to generate a signal spectrum centered at carrier third harmonic which will cancel the corresponding output of the carrier modulator. Successful translation of the video spectrum to the L3 carrier band in a single modulation step depends upon the maintenance of this and other modulator balances to unusually stringent requirements.

The carrier supply oscillators for the modulators and demodulators are of the Meacham bridge type⁵ with quartz crystal frequency control and thermistor amplitude control. Frequency stability of two parts per million is required for successful carrier regeneration at the receiver. A constant temperature oven for the quartz plus the inherent stability of the bridge type circuit is expected to provide the required frequency stability between monthly maintenance periods.

A feature of the signal transmitted over the L3 system is a degree of modulation which exceeds the value commonly referred to as 100 per cent modulation. The resulting waveform contains a maximum ratio of information to peak carrier, important from the standpoint of optimum signal to noise performance. Fig. 4 shows progressively the reduction in peak carrier amplitude which may be effected by subtraction of carrier component from a modulated signal. Figs. 4(b), (c) and (d) each contain the same amplitude of video modulation. Fig. 4(b) represents a video modulated carrier signal with maximum carrier occurring at tips of synchronizing pulses and a minimum carrier, equal to 20 per cent of maximum carrier, corresponding to picture white. Fig. 4(c) represents the same signal as Fig. 4(b) except that the 20 per cent excess carrier has been subtracted. This is the 100 per cent modulation case. Fig. 4(d) shows the effect of further carrier subtraction, (addition of negative carrier), to reduce to a minimum the peak amplitude of the modulated signal. The waveform of Fig. 4(d) employed for L3 transmission requires $7\frac{1}{2}$ db less maximum carrier power than that of Fig. 4(b) for the same transmitted information. The term "excess carrier ratio" has been devised to describe degrees of modulation which exceed 100 per cent. It is the ratio of peak carrier amplitude to the peak-to-peak modulation

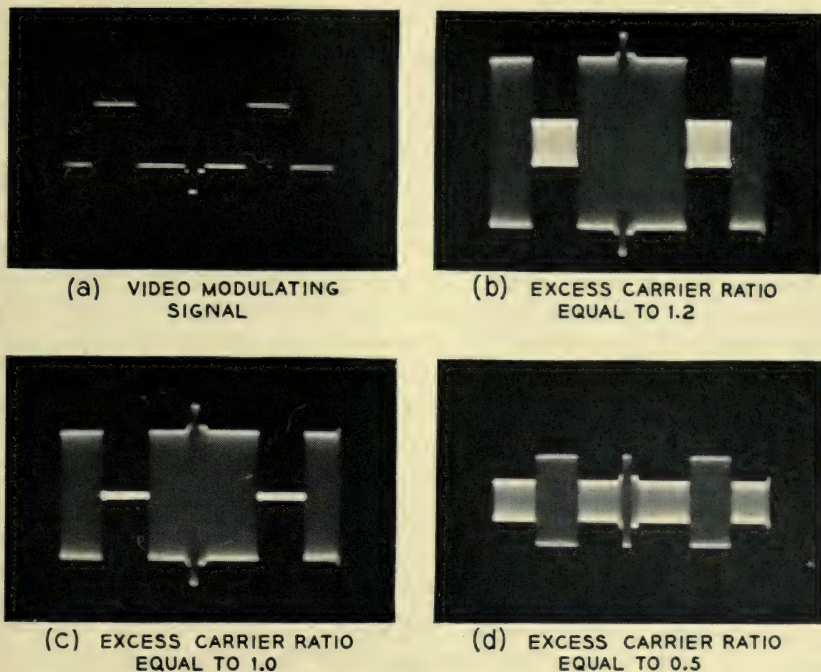


Fig. 4 — Carrier waves variously modulated by a composite video signal.

amplitude. Excess carrier ratio, (ECR), for the waveforms of Figs. 4(b), (c) and (d), respectively, are 1.2, 1.0 and 0.5.

Modulated signals of the forms of Fig. 4(b) or 4(c) may be detected by rectification, i.e., envelope detection. However, rectification of the waveform, Fig. 4(d), produces a spurious envelope wherein video signals which exceed a particular value are inverted. It is necessary to employ homodyne detection, that is, a demodulator driven by a locally generated carrier which is synchronous in phase angle and frequency with the carrier component of the signal wave. As described later, homodyne detection also makes possible the necessary suppression of the quadrature distortion associated with vestigial sideband transmission. Quadrature distortion associated with envelope detection is tolerated in the L1 coaxial system but with tandem operation of terminals required in the L3 system, would accumulate to intolerable values.

VESTIGIAL SIDEBAND CONSIDERATIONS

A vestigial sideband signal is produced by a band shaping filter following the modulator. In this filter the lower sideband is suppressed com-

pletely except for those frequencies which are within 500 kc of television carrier. Lower sideband frequencies within 500 kc of carrier are suppressed only partly as also are upper sideband frequencies within 500 kc of carrier to achieve a symmetrical response function in the vestigial sideband region.

It is convenient in a discussion of vestigial sideband transmission to consider the transmission as made up of two components, each symmetrical about carrier frequency, a real or in-phase component and a quadrature component which is a distortion term.⁶ The process is illustrated in Fig. 5. Here the response function shown in Fig. 5(a) represents idealized conditions for vestigial sideband transmission. The main sideband is shown extending from carrier frequency F_c to the upper cut-off F_u . A vestige of the lower sideband extends from carrier frequency to the lower cut-off F_v . Constant envelope delay is required in the entire band from F_v to F_u . In the frequency region $F_c \pm F_v$ the response characteristic is so shaped that the sum of responses at corresponding frequencies above and below carrier add to a constant value. The summing of signal components in the vestigial bands above and below carrier is accomplished by the receiver demodulator.

The response function of Fig. 5(a) may be considered to be the sum of the two response functions 5(b) and 5(c) which have even and odd symmetry respectively about the carrier F_c . Both 5(b) and 5(c) are double sideband functions. The component in Fig. 5(b) represents the

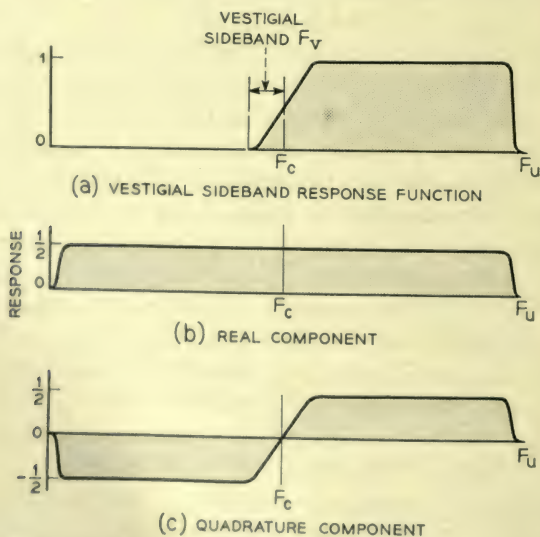


Fig. 5 — Response function of a vestigial sideband carrier system.

normal double sideband output of the modulator supplied with video and carrier signals. The other component, Fig. 5(c), represents the output of a modulator supplied with carrier and signal voltages each shifted in phase by 90° and with the amplitude attenuated as shown in the vestigial region $F_c \pm F_v$. The modulation transmitted by a circuit with the response function of Fig. 5(c) is called the quadrature component and is related to the normal modulation, depending upon the shape and extent of the vestigial sideband. Fig. 6 illustrates the real and quadrature components of an idealized rectangular wave form demodulated after transmission over circuits having the response functions of Fig. 5, respectively.

The 90° shift of carrier frequency in the quadrature component of the vestigial sideband signal makes possible the suppression of this component. The transmitted vestigial sideband signal may be written

$$V(t) = P(t) \cos ct + Q(t) \sin ct, \quad (1)$$

where $c = 2\pi$ times carrier frequency and $P(t)$ and $Q(t)$ are "real" and quadrature modulating functions⁶ typically as represented on Fig. 6.

The demodulator may be regarded as an ideal multiplier of signal and carrier supply. Let the carrier supply be denoted:

$$C(t) = \cos(ct - \phi), \quad (2)$$

where ϕ is the phase of the receiver carrier supply relative to the carrier factor of the "real" component of the signal. The demodulator output is the product $V(t) \times C(t)$. A low-pass filter rejects the output components in the band of frequencies about twice carrier frequency so that the demodulated video signal output is the lower frequency component. Thus, neglecting a factor of $\frac{1}{2}$,

$$V_0(t) = P(t) \cos \phi + Q(t) \sin \phi. \quad (3)$$

It is seen that real and quadrature video components in the demodulator

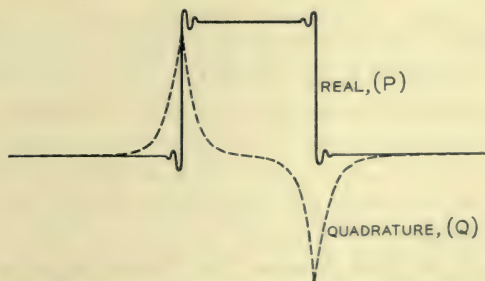


Fig. 6 — Real and quadrature components of a rectangular pulse after vestigial sideband transmission.

output exist in the same proportion as the components of the demodulator carrier supply in phase and in quadrature respectively with the real carrier component of the vestigial signal. By providing carrier exactly in phase with the real component of the signal the quadrature component in the output may be suppressed completely. It has been determined that to suppress the quadrature component resulting from the L3 vestigial band shape to barely perceptible (threshold) values the phase angle of the carrier regenerated at the receiver must be maintained to an accuracy of plus or minus 2.5 degrees. A requirement for one demodulator, when six pairs of terminals contribute to produce quadrature distortion at threshold value, becomes 2.5 degrees divided by the square root of six, or about one degree.

The regeneration of carrier at the receiver is one of the principal L3 terminal features. Here a 4.139-mc carrier must be provided to demodulate the "over-modulated" L3 signal. The required carrier must be reconstituted from information carried in the signal itself. It would be possible, of course, to transmit separately a signal from which carrier frequency could be derived but carrier frequency is really the smallest part of the required information. It is the phase angle of the carrier of the received signal which must be duplicated closely at the demodulator and separate transmission of carrier phase angle does not seem feasible. A phase controlled oscillator is employed for the carrier supply at the receiver, with phase control obtained from information residing in the signal itself and frequency synchronization an additional burden upon the phase control system.

The basis for synchronizing the receiver oscillator to the carrier of the received signal lies in the phase angle of the carrier frequency component of the vestigial sideband signal averaged over a period of time of the order of one frame scanning period. Referring to Fig. 5(b) and 5(c) again, it may be noticed that the quadrature response function is zero at carrier frequency. This means that the quadrature component of the transmitted signal contains no carrier frequency component and will not affect the determination of the real carrier component phase angle based upon averaging over a sufficient period of time. Another signal characteristic presents more serious problems. The degree of modulation employed in L3, shown in Fig. 4(d), makes the average carrier polarity indeterminate. That is, the carrier polarity for a video amplitude corresponding to picture white is opposite to that corresponding to picture black or sync pulses. The polarity reverses as the composite signal changes through its half peak-to-peak value. The average polarity determined from a predominantly white picture is thus opposite to that

determined from a predominantly black picture. A carrier oscillator, phase synchronized to the average carrier phase of the signal would execute 180° phase reversals as picture content changed from average white to average black, producing sudden video polarity reversals at the demodulator output.

This signal carrier polarity ambiguity which is momentary in character can be exchanged for one which is not time variable by a multiplication operation. The modulated signal is squared, i.e., multiplied by itself on an instantaneous basis, in a square law circuit. Such an operation squares carrier amplitude and doubles carrier frequency and phase angle, the latter effect converting 180° degree phase reversals into 360° degree changes which are indeterminable in the average phase detector. Under these conditions the phase synchronized demodulator carrier supply, stably locked to the average phase of the squared signal, experiences no phase reversals with change in picture content. The ambiguity now is in the determination of incoming signal polarity. The squaring operation eliminates any basis for determining polarity so that the demodulator carrier may with equal likelihood lock to either polarity relative to the signal and thereby at the demodulator output produce video signal waveforms of either polarity.

The method used to secure phase synchronization of the local receiver oscillator to the received signal is described next with reference to Fig. 7. Signal from the line together with the output from the carrier oscillator are brought to the demodulator where the desired video output signal is obtained as the lower sideband of the modulation product. This process has already been described, equations 1 to 3. In the carrier regeneration process signal and carrier phase shifted by 45° (equations 4 & 5) are each squared in square law circuits.

$$V(t) = P \cos ct + Q \sin ct, \quad (4)$$

$$C(t) \angle 45^\circ = \pm \cos (ct - \phi - \pi/4). \quad (5)$$

Band pass filters select from the squaring circuit output signal frequencies in the neighborhood of twice carrier frequency. From the signal squarer,

$$\frac{1}{2}(P^2 - Q^2) \cos 2ct + P Q \sin 2ct \quad (6)$$

and from the carrier squarer,

$$\frac{1}{2} \sin (2ct - 2\phi). \quad (7)$$

The two squared signals at twice carrier frequency are multiplied together in a product modulator. This product contains signals in two

bands of frequencies, one band in the region of four times carrier frequency and the other in the video frequency band starting at zero frequency. The lower frequency component of this product is selected by a low pass filter following the product modulator yielding,

$$\frac{(P^2 - Q^2)}{2} \sin 2\phi + PQ \cos 2\phi. \quad (8)$$

The dc component of the low pass filter output is a suitable control voltage for synchronization and is obtained in the limit as the cut-off frequency of the low pass filter is lowered. Average values as produced by the low pass filters are applied as a frequency control voltage to the carrier oscillator.

The first term in equation (8)

$$\frac{(P^2 - Q^2)}{2} \sin 2\phi$$

when averaged is, for small errors in carrier phase angle, proportional to the error angle ϕ . The factor of proportionality is recognized as the difference in mean squared values of the "real" and quadrature modulating functions, P and Q, illustrated typically in Fig. 5. This difference is always positive when the modulating signal contains energy components

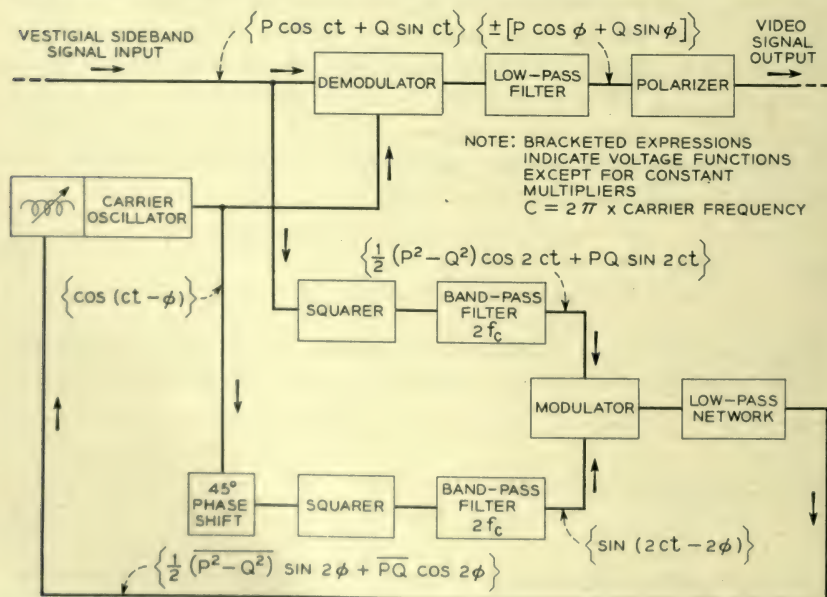


Fig. 7 — Functional diagram of the L3 homodyne demodulation process.

within the bounds of the vestigial sideband since the quadrature response function, Fig. 5(c), is attenuated relative to the "real" response function in this band. In the present case, with a 500 kilocycle vestigial bandwidth and a composite video waveform for a modulating function, the amplitude of Q^2 for control purposes is negligible compared with P^2 .

The second term of equation (8), $PQ \cos 2\phi$, contains no dc component since Q itself contains no dc component and all other frequency components of Q are shifted 90° in phase relative to corresponding components in P . The dc control voltage therefore is not modified by the existence of the second term of equation (8). However, the function $PQ \cos 2\phi$ does contain sum and difference frequencies due to the cross products of the spectra of P and Q . These frequencies in the control voltage tend to be large compared with corresponding frequencies due to the products P^2 and Q^2 since the trigonometric multiplier $\cos 2\phi$, equation (8), is large when the phase angle error is small. The effect of the term $PQ \cos 2\phi$ is that of phase modulation of the receiver carrier supply and its suppression determines the characteristics required of the low pass filter which averages the control signal. At the penalty of sluggish synchronization and restricted oscillator pull-in range the phase modulation can be reduced to arbitrarily small values. For our purposes a pull-in range of ± 20 cps can be achieved with phase modulation less than ± 0.1 degree with adequate margins.

The control voltage is applied to a tuning element in the receiver oscillator, in this case a small saturable reactor made with ferrite as a core material. This reactor is part of the series resonant quartz crystal circuit which determines the oscillator frequency and is capable of shifting the frequency in response to the control voltage by ± 20 cps, a figure chosen as safely less than the first sideband components of the transmitted signal which are ± 30 cycles from carrier frequency. This precaution avoids possible synchronization of the local oscillator to a signal sideband frequency rather than to the carrier.

Sufficient gain is provided in the carrier frequency control loop just described so that the maximum frequency difference encountered between transmitting and receiving oscillators is corrected by the phase control voltage due to a steady state phase angle error, ϕ , less than $\frac{1}{2}$ degree. The control characteristic of the saturable reactor may be expressed,

$$\Delta f = A(P^2 - Q^2) \sin 2\phi, \quad (9)$$

where Δf is the frequency shift introduced by the reactor and A is the factor proportional to required loop gain.

One other factor to be considered is the stability criterion of the

frequency control circuit as a feedback loop.⁷ In this case two factors contribute to loop phase shift. First, the phase angle variation of the oscillator output in response to the control voltage is an integration process. The control voltage changes the oscillator frequency and the resulting phase change can be expressed as the integral with respect to time of the frequency shift,

$$\text{phase, } \theta, = 2\pi \int \Delta f dt. \quad (10)$$

The integration with respect to time introduces a 90 degree "low-pass" phase shift into the control loop at all frequencies. Second, the averaging low-pass filter introduces phase shift in the same direction so that care must be exercised to avoid instability. In this case a phase stability margin of 45 degrees is provided over a wide range of frequency by designing the low-pass filter as a series of resistance capacitance steps of loss. These are staggered in frequency to produce a cut-off rate of 3 db per octave with a phase shift of 45 degrees over a wide frequency band.

The polarity ambiguity resulting from the squaring process has been demonstrated in the derivation of the phase control voltage. In equation (5) the plus or minus designation indicates that either polarity of carrier signal might be assumed without affecting subsequent expressions. However, in the derivation of the output voltage from the main signal demodulation, equation (3), the output signal polarity reverses if the carrier, equation (2), is assumed with reversed polarity. The carrier polarity established at any given time depends largely upon initial phase conditions when signal is applied.

Correct video polarity at the receiver output is established by a new device called a polarizer which follows the demodulator. This circuit recognized video polarity on the basis of standard features in the composite television waveform. The particular features used in this case are the vertical blanking discontinuities expected once each sixtieth of a second and the duty factor of sync pulses. These two characteristics taken together form a sufficient condition for the determination of polarity of any composite video waveform independent of picture content. The polarity, once recognized to be inverted, is corrected.

HARMONIC DISTORTION

A significant consideration in transmission problems is the generation of distortion by the non-linear amplitude characteristic of the transmission apparatus. In the case of video transmission, non-linearity results

in the distortion of brightness values of the transmitted picture and presumably will distort chromaticity values of color television signals. With carrier transmission apparatus, in addition to these effects non-linearity produces extraneous interference patterns at harmonics of the carrier frequency. In L3 with a carrier modulated spectrum concentrated near 4.139 mc., the second harmonic distortion of line repeating amplifiers produces a new spectrum concentrated near 8.278 mc, which is demodulated by the receiving terminal to the region near 4.139 mc. This form of distortion is considerably more disturbing in the final picture than a comparable distortion of brightness values and constitutes a limit to transmission signal-to-noise performance.

Advantage is taken of the spectral distribution of energy of television signals to ameliorate somewhat the effects of second harmonic distortion. A pre-emphasis network is employed in the transmitting terminal to accentuate the amplitude of high frequency components of the signal before transmission. At the receiver a restorer network introduces a complementary frequency characteristic to make the overall transmission characteristic constant with frequency, (see Fig. 8). The restorer network, de-emphasizing the high frequency components, likewise suppresses the second harmonic distortion signals. A limit to the amount of predistortion permitted is set by the maximum amplitudes expected of the high frequency picture components particularly in anticipation of a high frequency color sub-carrier in color television systems. Tentatively, the characteristic of Fig. 8 is chosen as a compromise of these factors.

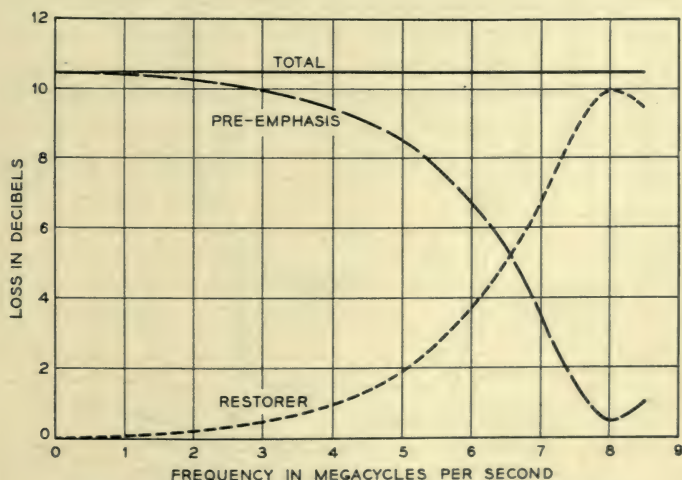


Fig. 8 — High frequency pre-emphasis characteristic.

FILTERS AND EQUALIZERS

A low-pass filter is required before the video signal is modulated, to limit the bandwidth of the signal, eliminating possible sources of disturbing cross products, both in the modulator and in the repeaters of the L3 system. This filter has a cut-off at 4.3 mc and provides over 40 db discrimination to all frequencies greater than 4.8 mc. It consists of three m -derived sections and introduces about 1.3 microseconds of envelope delay distortion near its cut-off frequency. This is equalized after modulation by the delay equalizer to be described later.

Following the modulator is the vestigial sideband filter which passes the upper sideband, and provides 60 db discrimination against all frequencies in the lower sideband less than 3.7 mc. The large discrimination is required to avoid interference with the telephone channels during transmission over the coaxial line. This filter provides a controlled loss characteristic to frequencies in the band 3.64 to 4.64 mc which satisfies the requirement for vestigial sideband transmission. The response function for this band is shown on Fig. 2. A flat transmission characteristic including the effective pass band loss of the video LP filter is maintained over the entire upper sideband from 4.6 to 8.44 mc. A second low pass filter after the modulator provides at least 50 db discrimination against third and higher harmonics of the TV carrier (4.139 mc).

A four section high-pass filter designed by the insertion loss method⁸ is used to supply 50 db of the discrimination at frequencies less than

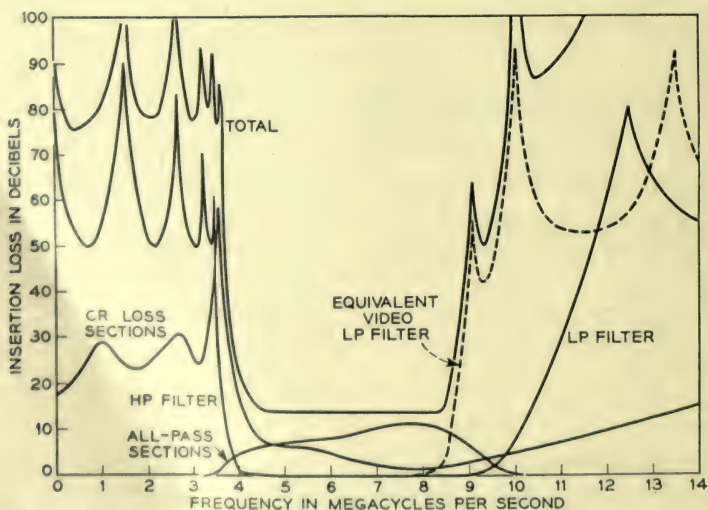


Fig. 9 — Insertion loss of the transmitting terminal filters.

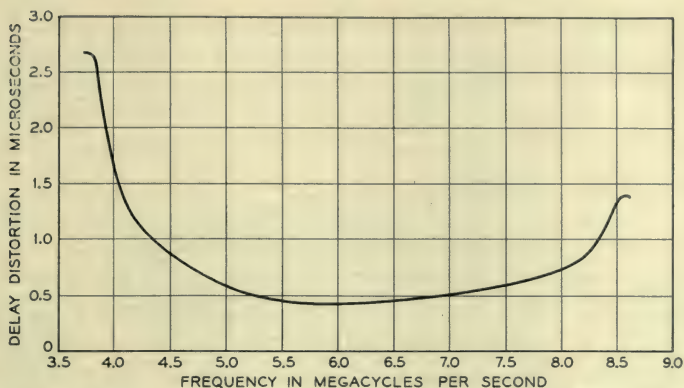


Fig. 10 — Envelope delay distortion of the transmitting terminal filters.

3.5 mc. This filter has a low-pass shunt network at the input to maintain its stopband impedance at 75 ohms. This is required because to produce a uniform frequency characteristic the impedance facing the modulator has to provide a reflection coefficient not exceeding 3 per cent. The remaining discrimination at frequencies less than 3.7 mc and the major part of the vestigial sideband shaping are provided by a group of six constant resistance equalizer sections, as shown on Fig. 9. The low-pass filter to suppress carrier harmonics consists of 2 m-derived filter sections.

The delay distortion of the complete set of filters and loss equalizer sections is shown on Fig. 10. This includes the equivalent delay distortion of the video low-pass filter, translated in frequency for equalization after the modulator. The distortion must be equalized to a constant delay over the television band. A delay equalizer to do this is incorporated in the filter. It was designed by a potential analogue method⁹ and consists of 24 all-pass sections, each having the schematic as shown in Fig. 11. A redundant capacitor is used to avoid excessively small capacity values.

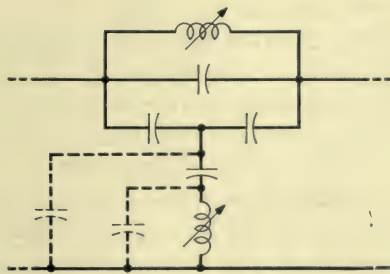


Fig. 11 — Schematic of a delay equalizer section, (dotted capacitors are parasitic).

The capacitors shown dotted are parasitic elements which must be compensated for by modifying the design values of the other elements. This group of sections has an insertion loss varying from 2 to 12 db across the pass band of the filter, caused by the dissipation of the elements. This loss, of course, must be taken into account in the design of the loss equalizer sections.

The objective is to obtain equivalent video transmission through the terminal flat to limits varying from ± 0.02 db to ± 0.10 db, depending upon the frequency characteristic of the deviation. To achieve this, close control of the dissipation loss is essential. The inductor losses which cause the major part of the delay equalizer loss characteristic vary up to ± 15 per cent from nominal values. As a means for controlling inductor Q to ± 0.5 per cent or better, a " Q adjusting screw" is employed. The inductors are solenoids wound on molded tubes with a threaded hole through the center. A threaded magnetic dust core is used for inductance adjustment. Its travel can be limited to the distance from the center to one end of the form without losing adjustment range. By introducing an additional core made of solid magnetic iron into the field of the solenoid, using the opposite end of the form, an adjustment is provided which reduces the Q in a continuous manner as the second screw is advanced into the form. The reduction in Q is caused by the losses in the iron, and normally these would cause a reduction in inductance also. However, the permeability of iron causes an increased concentration of field which tends to increase the inductance. A balance between these tendencies to decrease and to increase the inductance is obtained by controlling the geometry of the Q adjusting core. As a result, a reduction of up to 50 per cent in Q can be obtained, accompanied by a change of less than one per cent in the inductance. Models of the inductor and the adjusting screws are shown on Fig. 12. This adjusting screw in conjunction with the magnetic dust core provides an accurate and economical means for adjusting simultaneously both inductance and dissipation in each inductor of the delay equalizer.

The flat transmission level for the upper sideband and the shaped cut-off for the vestigial sideband were obtained by including loss equalizer sections, assuming Q factors for the all-pass sections of about 20 per cent less than the nominal Q of the inductors. As a final step in the design, the Q factors were modified to absorb in the loss of the all-pass sections the residual loss distortion uncompensated by the loss equalizer sections. This in effect provided the use of 24 additional parameters for shaping the loss in the pass band and resulted in an improved loss characteristic.

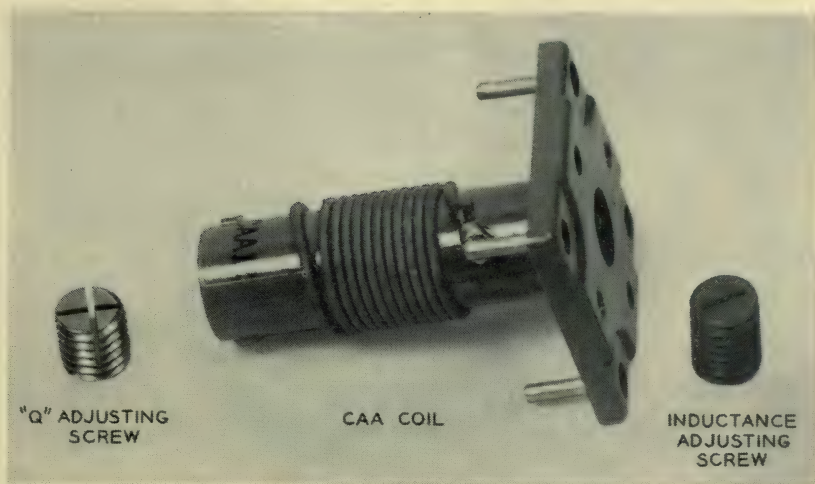


Fig. 12 — Inductor with adjusting screws.

The close limits on delay distortion can be met only by close control of the adjustments on the individual all-pass sections. In order to obtain reproducible results to the order of $\pm 0.1^\circ$ for the phase shift and ± 0.01 db for the insertion loss of the individual sections, a special fixture is employed to make the connection between the section and the measuring circuit. This is shown in Fig. 13. The fixture can be clamped on the network terminals quickly without soldering and provides coaxial patch cords with plugs for connection to the measuring circuit. Each section is mounted in an individual container with shielding between the inductors to reduce coupling. Each inductor is resonated with its associated capacitors and the dissipation is adjusted by adjusting the two cores. The construction of a typical section is illustrated by Fig. 14.

An important consideration in obtaining smooth loss and delay characteristics for the delay equalizer is the reflection coefficient of each section. Poor reflection coefficients cause reflections and interactions between all-pass sections. Due to the large phase slope of the equalizer these tend to produce frequency characteristics with large numbers of loss and delay ripples across the frequency band for which transmission requirements are most severe. Reflection coefficients of 2 per cent or less at all frequencies in the TV band have been obtained for all delay sections by taking the following precautions:

1. Mutual coupling is limited between the two inductors in each section by use of a shield in the section container. As little as 0.1 per cent

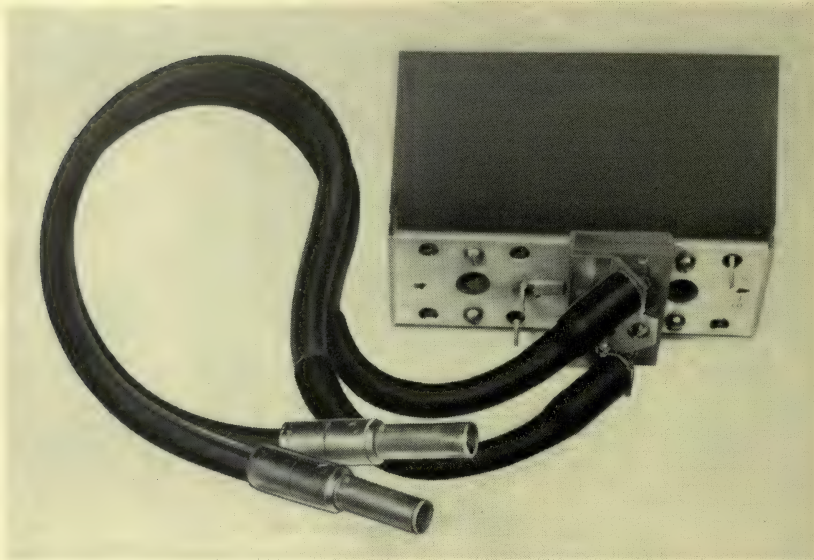


Fig. 13 — Fixture for delay equalizer section adjustment.

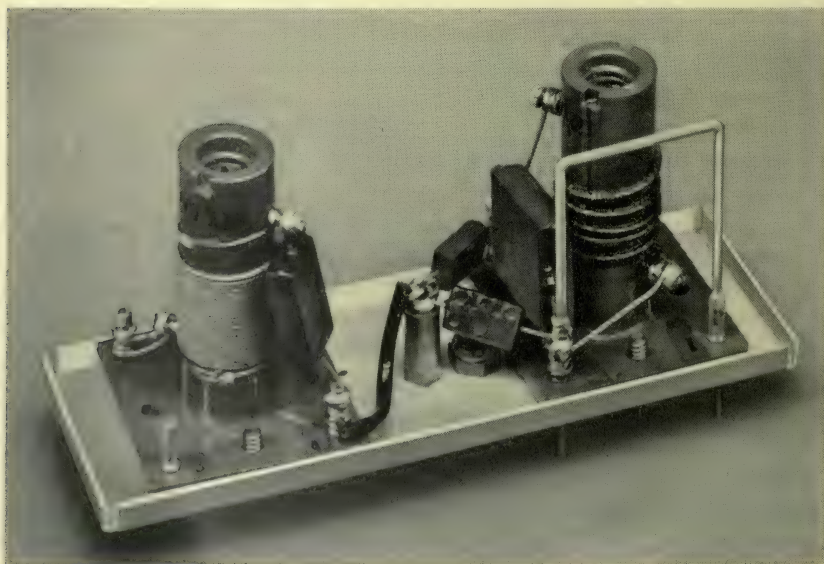


Fig. 14 — Model of a typical delay equalizer section.

coupling coefficient can cause a reflection coefficient of 1 per cent in certain sections at the frequency of 180° phase shift.

2. The Q factors of both inductors are adjusted to be equal in each section.

3. The values of the capacitors are modified to compensate for the presence of the parasitic capacitances associated with the shunt arm, Fig. 11.

The measured insertion loss characteristic and phase shift deviation from linear phase slope for one model of the transmitting filter and delay equalizer is shown on Fig. 15. This has been reduced to video frequency to show the detailed residual distortion which results from the addition of the vestigial lower and upper sideband. The delay equalization is maintained for about 200 kc above the loss cut-off to provide for at least 30-db insertion loss at frequencies where the delay distortion becomes large. Without this precaution the transient response is characterized by a severe "cut-off ring" distortion which is a slowly damped oscillation at cut-off frequency generated by high frequency signal components.

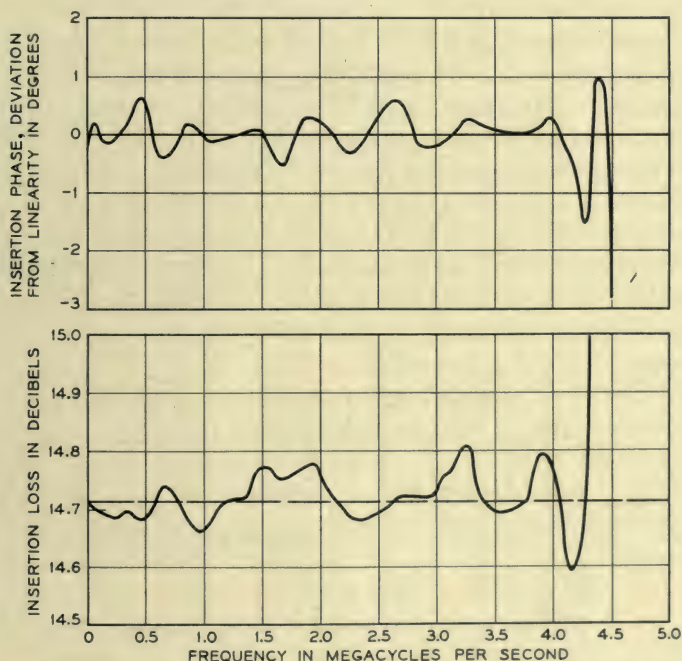


Fig. 15 — Measured pass band performance of a trial model of the transmitting filters and equalizer.

Variations in transmission response in the region near television carrier cause noticeable "smear" distortion on the received images. These variations can be introduced by small changes in element values of the filter or equalizer sections following initial adjustment, or by other changes in both transmitter and receiver. As a means of compensating for such variations, two adjustable loss equalizers have been provided, each with adjustable maximum loss at carrier frequency. One has a half-loss point at 300 kc and the other at 80 kc from the carrier. These are adjusted for minimum "smear" at periodic intervals.

A fixed equalizer is provided in each terminal to compensate for loss and delay distortion other than that in the filters described above. For convenience, this equalization is done at modulated frequencies. It is expected that additional means for periodic re-equalization of the terminal gain and phase characteristics will be necessary to achieve 4,000 mile, tandem terminal transmission quality objectives.

PILOT FREQUENCIES FILTERS

As described elsewhere,¹ the L3 system has six pilot frequencies for regulating automatically the transmission characteristic of the line. Two of these, 7.266 mc and 8.320 mc, are in the television band. At the transmitting terminal a pilot elimination filter for these two frequencies is required to prevent energy in the TV signal near these frequencies from disturbing the pilot levels on the system. At the receiving terminals, these pilot frequencies must be removed from the TV band to avoid interfering effects in the output TV signal. A discrimination of 50 db for frequencies within 20 cycles of the two pilot frequencies is required. Also at the receiver a carrier elimination filter is required to suppress the residual carrier leak from the demodulator. This requires 30 db suppression to frequencies within 20 cycles of 4.139 mc.

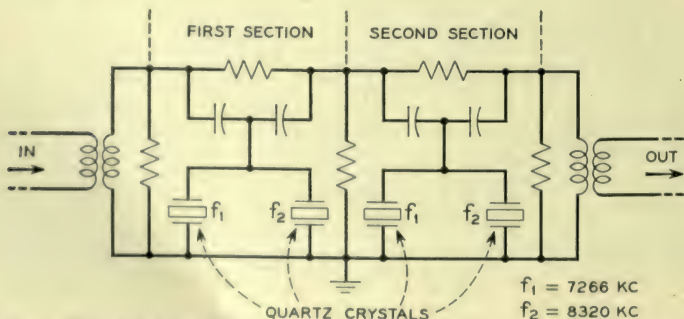


Fig. 16 — Simplified schematic of the pilot frequency band elimination filter.

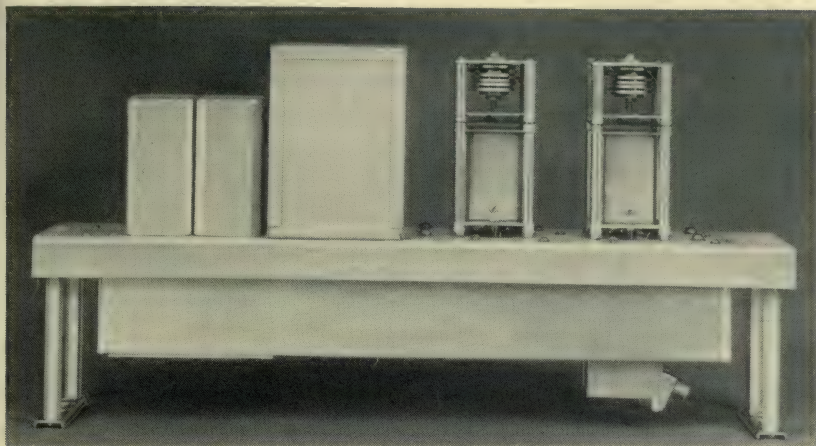


Fig. 17 — Model of the carrier suppression filter.

To avoid removing excessively broad bands of frequencies from the television band, and thereby generating visible distortion in the transmitted picture, narrow bandwidth crystal filters are used for both of these applications. The pilot elimination filter has its 3 db loss points at about 1,000 cycles on either side of the pilots. The carrier suppression filter has its 3 db loss points at 150 cycles on either side of the carrier. In the past, spurious or secondary responses in the crystal units could not be limited to sufficiently low values to permit available crystals to be used in broad band circuits without elaborate means for suppressing the unwanted responses. In the L1 system, for example, a balancing circuit using hybrid coils was employed for this function. Techniques for reducing unwanted responses in the crystal units by special contouring of the blanks and by precise optimum area plating have been developed recently.¹⁰ The use of these methods has resulted in crystal units in which the unwanted responses are reduced to the extent that relatively simple filters can be used for these applications. A simplified schematic of the pilot elimination filter is shown on Fig. 16. Small ovens are provided to maintain close temperature control for the quartz crystal elements in order to stabilize the resonant frequencies. Mechanical arrangements are illustrated on Fig. 17.

POWER EQUIPMENT

Primary power for the television terminals is 60-cycle ac. It is derived from the L3 motor-alternator power equipment at main repeater sta-

tions or, alternatively, from commercial 60-cycle sources. When commercial power is used directly a provision is made for switch-over to emergency power in the event of power failure. The emergency power supply comprises a 130-volt battery, an inverter, and automatic switch-over equipment. If during service the commercial power source fails switch-over to emergency power is made automatically in a time under one-tenth second.

Dc and heater power provisions involve two departures from previous practice. First, the rectifiers which supply anode and other dc power requirements are non-regulated metallic, (selenium), rectifiers. These are supplied from magnetic ac line voltage regulators to obtain suppression of line voltage variations. The combination of the magnetic ac regulator and the metallic rectifier provides adequate suppression of hum and line voltage variations with, it is expected, greater reliability than alternative vacuum tube regulated rectifiers. Secondly, the magnetic ac line voltage regulator together with an improved heater transformer design make possible the operation of the vacuum tubes at reduced heater voltage to obtain increased thermionic life expectancy. This advantage can be taken only if close control of heater transformer output voltage is possible. In this case, variations in output voltage expected to occur due to temperature variations of the transformer windings substantially have been eliminated by incorporating thermistor temperature compensation in the transformer primary circuit.

T3 TRANSMITTING TERMINAL

A block diagram of the components of the transmitting terminal is shown in Fig. 18. The composite video signal is received from the Bell System television operation center over 124-ohm balanced cable and

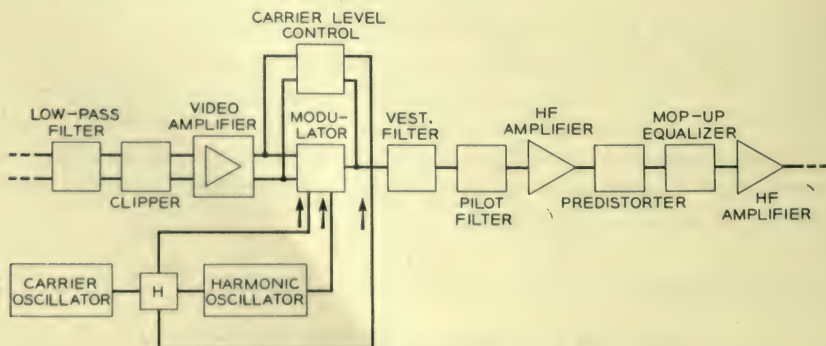


Fig. 18 — L3 transmitting terminal block diagram.

the modulated output signal is delivered to the L3 line facilities over 75-ohm coaxial cable.

An input low-pass filter and a video clipper circuit place ceilings on the maximum transmitted signal bandwidth and signal amplitude, respectively. The clipper circuit is required to protect the 600 telephone channels from inadvertent overloads due to excessive television signals. Normal amplitude signals are not affected by the clipper.

Following are a video amplifier and the modulator together with the required carrier supplies and the carrier level control. This latter device accurately regulates the peak carrier magnitude in the output signal and thereby preserves the desirable maximum modulation.

Following the modulator is the vestigial band filter which, together with its delay equalizer, shapes the double sideband modulator output into the vestigial transmission band. The flat loss of these networks requires that amplification be provided in the carrier frequency band to increase the signal level. The first of two high frequency flat gain amplifiers in the transmitter restores the signal amplitude.

The pilot band elimination filter is next provided to remove television signal energy and other possible interference from particular frequencies allocated to the L3 line pilot signals.

At this point occur the predistorter network for pre-emphasizing the high frequency signal components and a mop-up equalizer. The mop-up equalizer is to provide means for periodic correcting of the transmission characteristic. The high-frequency amplifiers, particularly, change their transmission characteristic as the vacuum tubes age. An additional high frequency amplifier is provided to recover the loss of the foregoing networks and deliver a proper signal level to the line equipment.

R3 RECEIVING TERMINAL

The receiver demodulates the signal as transmitted over the L3 line facilities to recover the video signal for transmission over local video circuits. As has been discussed the demodulation is a homodyne process utilizing a local carrier regenerated from information contained in the transmitted signal.

A block diagram of receiver components is shown in Fig. 19. The first components are a group of networks; the restorer to compensate for transmitter predistortion, the fixed and variable mop-up equalizers to compensate for deviations in the receiver frequency characteristic, and a pilot elimination filter to remove L3 pilot frequencies from the signal. A high frequency amplifier provides amplification to compensate for the network losses.

The amplified signal is split into two branches by a hybrid transformer. Part of the signal is carried to the demodulator for detection to video frequencies while the remainder is employed in the carrier regeneration apparatus. The carrier supply oscillator output likewise is split between the same two circuits; the portion supplied to the demodulator providing the carrier energy for demodulation and the part sent to the carrier regeneration circuits providing means for comparison with the input signal. The carrier regeneration equipment comprises mainly the phase comparison circuit together with means for changing electronically the frequency of the carrier supply oscillator.

The phase synchronized carrier supply together with the associated third harmonic supply provide demodulating carriers which translate the carrier spectrum down in frequency to the original video spectrum. The demodulator output also contains an unwanted upper sideband in the 8 to 12-mc region. This signal and residual carrier leak from the demodulator are removed from the detected signal by a low-pass filter and the quartz crystal narrow band rejection filter tuned at carrier frequency. Finally the signal is amplified at video frequencies to provide moderately high video frequency transmission levels and transmitted through the reversing relays of the polarizer to the receiver output.

Fig. 20 is a photograph of two equipment bays each eleven feet high which comprise at the left the T3 transmitting terminal and at the right the R3 receiving terminal.

Acknowledgement is made for the many contributions made by our associates toward the successful development of these terminals. These include both ideas related to transmission processes and important contributions in the development of many new circuit components which were necessary to the achievement of the quality and reliability objectives.

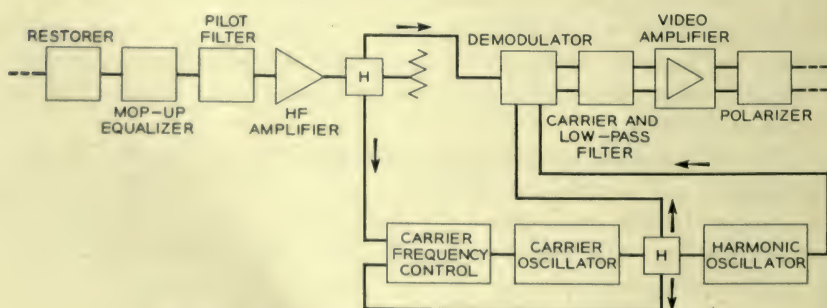


Fig. 19 — L3 receiving terminal block diagram.

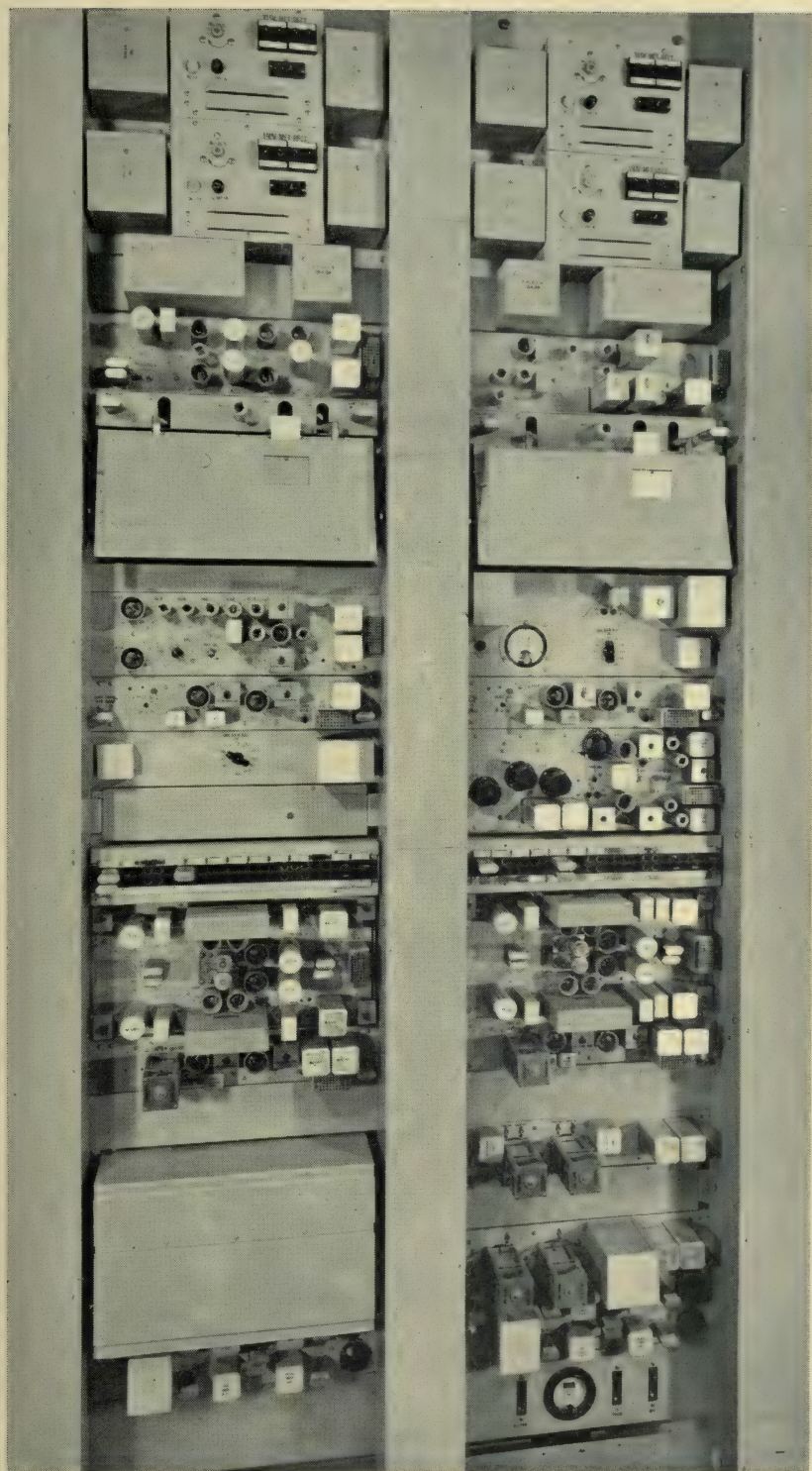


Fig. 20 — Model of transmitting and receiving terminal equipments.

REFERENCES

1. C. H. Elmendorf, R. D. Ehrbar, R. H. Klie, A. J. Grossman, The L3 Coaxial System—System Design, see pp. 781 to 832 of this issue.
2. L. H. Morris, G. H. Lovell, F. R. Dickinson, The L3 Coaxial System-Amplifiers, see pp. 879 to 914 of this issue.
3. L. W. Morrison, Jr., Television Terminals for Coaxial Systems, A.I.E.E., Trans. **68**, pp. 1193–1199, 1949.
4. R. S. Caruthers, Copper Oxide Modulators in Carrier Telephone Systems, Bell Sys. Tech. J., **18**, pp. 315–337, April, 1939.
5. L. A. Meacham, Bridge-Stabilized Oscillator, Bell Sys. Tech. J., **17**, pp. 574–591, Oct., 1938.
6. H. Nyquist, Certain Topics in Telegraph Transmission Theory, A.I.E.E., Trans. **47**, pp. 617–644, 1928.
7. H. W. Bode, Network Analysis and Feedback Amplifier Design, D. Van Nostrand Co., 1945.
8. S. Darlington, Synthesis of Reactance Four-Poles, J. of Math. and Phy., **18**, pp. 257–353, Sept. 1939.
9. S. Darlington, The Potential Analogue Method of Network Synthesis, Bell Sys. Tech. J., **30**, pp. 315–365, April, 1951.
10. L. J. LaBrie and D. F. Ciccolella, High Frequency Crystal Units for Use in Selective Networks, submitted for publication, I.R.E. Trans. (Presented at I.R.E. subcommittee meeting, Washington, D. C., Dec. 4, 1952).

The L3 Coaxial System

Quality Control Requirements

By H. F. DODGE, B. J. KINSBURG and M. K. KRUGER

(Manuscript received May 5, 1953)

Economic solution of the equalization problem in the L3 system required limitation of the excursions of the transmission characteristics from the design center. To implement this, a pattern of distribution requirements for component elements of the system was worked out utilizing basic quality control techniques. An analysis of the several methods employed is presented with particular emphasis on the intent of the choices which were made and the operating characteristics of the resulting procedures. One of the novel features is the three-cell selection method which insures that the product delivered has the kind of distribution that is wanted even while the process is in trouble distribution-wise.

1.0 REASON FOR REQUIREMENTS

1.1 EQUALIZATION PROBLEMS

The L3 system is a long distance carrier system designed to transmit either 1,860 telephone channels or 600 telephone channels plus one television channel. The band width is approximately 8 mc. The repeaters are spaced about 4 miles apart and in a 4,000-mile route, counting both line and office amplifiers there will be about 1,200 amplifiers in tandem.

There are two equalization problems. The first is equalization proper, i.e., delivery of a satisfactory signal to the customer from the transmission characteristic point of view. The television equalization design objective of the system is to meet a signal-to-echo ratio of 40 db. This corresponds to a uniform sinusoidal ripple of 1/10 db or a complex deviation pattern several times larger in amplitude. The telephone equalization objective is more lenient and allows deviations as large as 1 to 2.5 db, depending on length of circuit and number of links in tandem. Though care was exercised to make basic design decisions which would tend to ease the equalization problem, still the reduction of the accumu-

lated deviation arising in 1,200 amplifiers to a few tenths of a db is a formidable problem.

The second problem is concerned with the effect of equalization on the signal-to-noise performance. As deviations creep into a repeatered circuit, the transmission levels deviate more and more from the normal or design levels. This effect is lumped in one term, misalignment. The result of misalignment is degradation in signal-to-noise performance. Periodic equalization helps to limit misalignment in the succeeding repeater sections but does not eliminate the increase in noise or modulation which has occurred in the preceding repeaters. Thus, the objective is not only to equalize the over-all circuit, but also to keep the deviations all along the transmission line within the specified bounds in order not to exceed signal-to-noise margins.

1.2 NEED FOR CONTROLLING DEVIATIONS AT THEIR SOURCE

Let us examine what accounts for the magnitude of the gain-versus-frequency deviations arising in any one repeater. Let us assume that a particular element deviates +1 per cent from the design objective. Is this good or bad? The answer to this question may be had only if a deviation study of the repeater is made and its sensitivity to the deviation of the element under consideration is ascertained. There are two factors which contribute to the equalization problem: (a) the deviation sensitivity of the repeater to a given deviation of the element from its prescribed value, and (b) the actual deviation of the element itself due to all causes, including manufacture, temperature, aging, etc. To simplify our terms, factor (a) will be called the sensitivity of the element, and factor (b), the deviation of the element. The sensitivity is a function only of the circuit design and is independent of the performance of the individual element. The deviation of the element is a function only of its design and manufacture.

1.3 USE OF STATISTICAL QUALITY CONTROL METHODS TO ASSURE A CONTROLLED DISTRIBUTION

When the design of the L3 system was initiated, it was realized that the effect of the variability of the component elements could be materially reduced by the application of statistical quality control techniques to design and manufacture. Once a circuit design is available and a deviation sensitivity study is made, it is comparatively easy to formulate the desired limits on the variability of components. Actually the process of arriving at an individual tolerance objective is more complex

since there is a large area of give and take between the circuit and the component element designers.

Let us assume that the process of arriving at a satisfactory component element design has been completed and that the spread of the manufacturing limits has been set at $\pm y$ per cent. How will the gain deviations due to this element grow, as more and more repeaters, each containing one unit of this element, are placed in tandem? It is evident that the cumulative magnitude of gain deviations will depend on the distribution pattern describing the departure of individual units of this element from the prescribed value.

If all units of this element have a systematic deviation (equal in magnitude and sign) from the prescribed value, the cumulative gain deviation will be

$$n\alpha_1 \text{ db}$$

where n = number of repeaters in tandem, and

α_1 = gain deviation in one repeater due to y_1 per cent deviation of the element.

If the units of this element have a Normal distribution whose average coincides with the prescribed value, and whose extreme limits (say 3-sigma limits) are $\pm y_2$ per cent, then the averages of random groups of elements in n repeaters will be described by another Normal distribution, the corresponding limits of which will be $\pm y_2/\sqrt{n}$ per cent. Thus the over-all limits of gain deviation for n repeaters will be

$$\pm \frac{n\alpha_2}{\sqrt{n}} \text{ db} = \sqrt{n}\alpha_2 \text{ db}$$

If, however, the average of the distribution of individual units does not coincide with the prescribed value but is displaced by y_1 per cent, then the over-all gain deviation limits for n repeaters in tandem will be

$$(n\alpha_1 \pm \sqrt{n}\alpha_2) \text{ db.}$$

Thus, it is of first importance that the average of the individual units be controlled as closely as possible to the prescribed aimed-at value.

Of course, distributions other than Normal are possible. However, it is sufficient for most practical purposes to consider only the Normal distribution since a combination of a large number of distributions, which individually are not Normal, will tend to approximate a Normal distribution. This assumption is a reasonable one to make because of the large* number of different elements which are used to make up an L3

* There are over 100 component elements in the L3 amplifier of which about 12 have large element sensitivities and are therefore critical in evaluating performance. There are over 30 additional elements the sensitivities of which are also sufficiently large to require application of distribution requirements.

repeater. It should be kept in mind that this assumption would not be valid if the contribution of one element to the deviation pattern of the repeater becomes dominant, say, larger than all the other elements combined.

In the L3 system, the first stages of equalization are spaced about 25 repeaters apart. In accordance with the above considerations, a distribution of individual element variations, the average of which is controlled close to the nominal, will result in cumulative gain deviation which will be substantially smaller than if no restrictions were placed on the average. Thus, a desired objective for critical L3 component elements is to provide a stabilized production process giving a distribution of individual values (a) having an *average* that is maintained consistently close to a desired nominal, and (b) having a pattern of variation around the average that is Normal (or nearly so). If this is attained, comparatively wide limits for the individual units are acceptable. Furthermore, assembly of component elements into amplifiers can be made on a random basis, and the problem of maintaining equipment in the face of replacement of parts failing in service will be greatly simplified.

From the very nature of the over-all problem the best approach to this objective has appeared to be through the application of statistical quality control methods, both in the design of and in the production of the component elements that are important from an equalization point of view.

2.0 BASIC FEATURES OF DISTRIBUTION REQUIREMENTS

2.1 GENERAL PLAN

For each of the important component elements, then, interest centers on closely controlling the collective quality of the product, especially the average of the individual values. This can hardly be accomplished merely by specifying and securing compliance with the usual type of requirements, expressed as maximum and minimum limits for individual units. Something more is needed. Consideration must be given to ways and means of placing requirements on the distribution of individual values from the successive increments of the product turned out day after day.

Accordingly, a general plan using quality control methods has been developed, specific features of which will be discussed in this paper with particular emphasis on the intent of certain choices that were made and on the procedures selected to meet the general objective. Further development work on some of these features may of course be found warranted as experience with them is gained.

The general plan has been implemented by imposing on important component elements certain so-called "distribution requirements" which incorporate quality control procedures for assuring a high degree of statistical uniformity in the quality of product delivered for service. The aim of the distribution requirements is to place a continuing limitation on the pattern and the spread of measured values (of a final critical characteristic of the product) around their average and to impose close limits on the departures of the average from a desired nominal value. To obtain these ends, close cooperation between the element designer and the production engineer is essential. In fact, compatibility of the specification requirements and the process capability is one of the basic provisions of the general plan. This should be established, if at all possible, in the design stage.

In some cases where distribution requirements have been applied to a final characteristic of an element, the production engineer has found it advantageous to introduce quality control techniques in some of the earlier manufacturing steps, as for example, on materials, piece-parts, or process operations which are found to have a major effect on end quality.* The character of such controls can rarely be planned in advance, but must be tailor-made to fit the particular process being used. Often too, a major difficulty encountered has been not the process itself but the precision and accuracy of the measuring equipment. The resolving of such problems during the design and the early production stages has been one of the aims of the general plan.

2.2 SELECTION OF CHARACTERISTICS TO BE CONTROLLED

The general procedure calls for imposing distribution requirements on not more than one characteristic of any component element. Before assigning limits on an individual component, therefore, it is important that the characteristic selected for control be the key characteristic. This statement seems to be trite, but its importance cannot be over-emphasized. Controlling all characteristics of a component is not only inherently uneconomical but may be found impossible in practice. In the case of an inductor, for example, it should be ascertained which characteristic is important to the circuit designer. It may be the value of inductance, of Q , of temperature coefficient, or of parasitic capacity. In addition, of course, requirements should be specific and apply, for instance, at a given frequency or within a definite temperature range.

* R. F. Garrett, T. L. Tuffnell and R. A. Waddell, The L3 Coaxial System — Application of Quality Control Requirements in the Manufacture of Components, see pp. 969-1006 of this issue.

Where it is desirable to exercise some control over one or more characteristics in addition to the key characteristic, a procedure is provided which requires that control charts be maintained on the additional characteristics. This procedure does not require a controlled distribution for such characteristics, but it does give a statistical record which shows the dynamic behavior of the process and indicates when remedial action is desirable.

2.3 COMPATIBILITY OF SPECIFICATION REQUIREMENTS AND MANUFACTURING PROCESSES

As mentioned above, special effort has been made in the L3 project to provide specification limits and manufacturing processes for individual component elements that are mutually compatible.

In order to determine realistic limits on the value of a particular quality characteristic, it is necessary to collect a reasonable quantity of data from the proposed process to show what it can do if brought into a state of statistical control. This is an area in which close cooperation between the element designer and the production engineer is necessary. Here it is convenient to define the "natural tolerance" of a process as the extreme range of variation to be expected among individual units of product made in relatively short periods of time, such as in single batches or production lots; mathematically it is taken to be equal to 6σ , where σ is the basic standard deviation of the process as estimated from the average spread for a *series* of samples, each selected from a different segment of production.

If it is found, for instance, that the natural tolerance of the process (6σ) is wider than the expected or desired specified limits, then a fundamental change either of the process or of the basic design of the component or both is called for, if mutual compatibility is to be attained. Of course, one way to avoid a major change would be to widen the specification limits. If the needs of the system, however, demand the closer limits, such a simple solution is not possible, and a manufacturing process or design change must be made. In many cases examined in connection with the L3 repeater, the economics of the situation — balancing the component cost against the saving in equalization gear — justified additional effort to improve designs and manufacturing processes to obtain limits for individual component elements narrower than those which initial processes appeared capable of meeting.

2.4 THE PROBLEM OF MEASUREMENT

The precision of measurement may have an important influence on the determination of the process capability. Let us assume that the

universe of true values for an element may be described by Curve A in Fig. 1, having a spread of $2r$. Let us also assume the distribution of errors of measurement representing the precision of the measuring device (assumed to be unbiased) is described by Curve B of Fig. 1, having a spread of $2s$. If, for the purposes of discussion, these distributions are Normal, the resulting apparent distribution of the individual values (the distribution of measured values) is a composite of the distribution A and B, as shown by Curve C of Fig. 1. The spread of this distribution will be $2q$, where $q = \sqrt{r^2 + s^2}$. If $s = \frac{1}{2}r$, as shown in Fig. 1, then $q = \sqrt{r^2 + 0.25r^2}$ or $q = 1.118r$. Thus, the apparent distribution has a spread which is about 12 per cent greater than the true distribution. Similar computations for other values of measuring precision in relation to the true distribution give the following:

Ratio s/r	Ratio q/r
1	1.414
0.5	1.118
0.2	1.020
0.1	1.005

Measurements normally made to determine process capabilities include the effect of random errors but not necessarily of systematic errors.

The effect of systematic error or bias of the measurements is quite different. Bias tends to cause unknown and unwanted displacement of the process average from the aimed-at value. This, in turn, can be re-

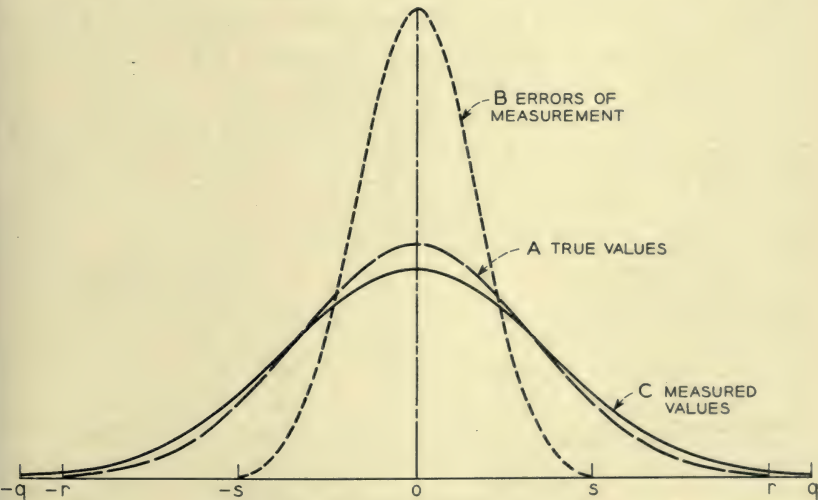


Fig. 1 — Effect of measuring errors.

sponsible for systematic deviations which are so undesirable in the L3 system. One way to combat or, rather, circumvent large bias is to rely on comparison standards in the calibration of test sets. The task of furnishing and maintaining reliable comparison standards has many pitfalls. It is an art or perhaps a science in itself and is mentioned here only because of its importance to the objective of restricting variations in the process average.

Considering the effect of both random and systematic errors of measurement, it has been found essential to place comparatively tight requirements on the accuracy of the test methods to be used. In many instances meeting these accuracy requirements was made possible by the development of new or radically improved measuring equipment for both laboratory and production purposes.

2.5 ALLOWABLE MARGIN FOR DRIFT OF PROCESS AVERAGE

Having determined σ , the basic standard deviation of an acceptable process for a quality characteristic, then the minimum spread of specified limits, to be compatible with the process, would be $\pm 3\sigma$ around the nominal (N). However, product having this σ could be expected to meet such limits practically all the time provided only that the average of the process were controlled at the nominal. Accordingly, provision was made to allow the process average itself to vary within a band of $\pm \frac{1}{3}\sigma$ around the nominal. This allowance is somewhat arbitrary and represents an estimate of relative importance to the L3 system of systematic changes

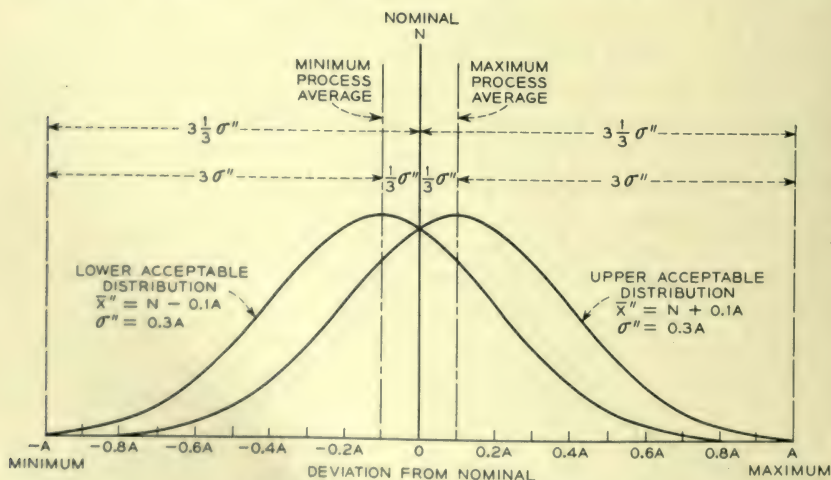


Fig. 2 — Basis of specified limits.

in the process average. Thus, two limiting acceptable distributions, Normal in shape, are defined as indicated in Fig. 2, both having a standard deviation, σ'' , equal to the above-mentioned basic standard deviation of the process, one having an average, $\bar{X}''$, located $\frac{1}{3}\sigma''$ below nominal and the other having an average $\frac{1}{3}\sigma''$ above nominal. The over-all limits for individual units then become $N \pm 3\frac{1}{3}\sigma''$. If we designate the distance from the nominal to either limit by A , then the permissible variation in the process average is $\pm 0.1A$.

Our discussions here will be confined to the case of characteristics having substantially Normal distributions and having both maximum and minimum specified limits. In the over-all plan provisions have also been made for characteristics having skew distributions or having a single specified limit, maximum or minimum.

2.6 EXPRESSION OF DISTRIBUTION REQUIREMENTS

With the decision to use distribution requirements, the question arose as to how such requirements should be expressed. Would it be adequate to write:

"The distribution of the individual values of characteristic X of the product shall meet the requirements: (1) average, not outside $(N \pm a)$, and (2) standard deviation, not greater than b ."? Should a clause be added saying that characteristic X "shall be statistically controlled?"

Recognizing that a requirement is not what is written but what it is interpreted to be, it appeared that such an expression or equivalent is not sufficient to explain the intent nor does it provide criteria for determining when any segment of production may be considered conforming to the intent. What is meant by "the product?" Does the limiting value on the "average" (or the "standard deviation") apply to each production segment of 10 units, each 50 units, each day's production, each six month's production? Can sampling be used? What criteria are to be used for judging whether the distribution of quality of a portion of product is satisfactory? What should be done with the product if it does not meet the criteria?

As with any requirement on the collective quality of an aggregation of like articles comprising a product, the use of distribution requirements brings in special problems on the clarification of the intent. What is really wanted is a flow of product such as would be obtained in a series of random samples of units from a process whose average is continually maintained within stated limits $(N \pm 0.1A)$ and whose standard deviation does not exceed $0.3A$.

With this in mind, the product specification prescribes for any given key characteristic a nominal value, N , maximum and minimum limits for individual units, $N \pm A$, and adds the clause "subject to the distribution requirements of (Y)." This supplementary specification (Y) gives a statistical description of the intent, together with control procedures to be followed in the inspection of the product.

3.0 CONTROL PROCEDURES ASSOCIATED WITH DISTRIBUTION REQUIREMENTS

Distribution requirements can be given operationally definite meaning by providing procedures (of the nature of inspection procedures) that define (a) the character and quantity of evidence needed regarding the collective quality of the product as it is made day by day, (b) the criteria for judging when such product may be considered conforming to the intent of the specification, and (c) the treatment or disposition of product units when these criteria are not met.

Three such procedures have been prepared to meet the several conditions that may be encountered in the production of L3 component elements:

1. control chart method,
2. batch method, and
3. three-cell method.

The first two methods permit the use of sampling. For both of these methods the criteria have been so selected that product should be found to be conforming to the distribution requirements practically all of the time if the process is so controlled as to maintain a distribution of individual units with (a) an average within the band, $N \pm 0.1A$, and (b) a standard deviation not greater than $0.3A$. The third method requires 100 per cent inspection, and while this method may be used at any time at the option of the manufacturer, its use is mandatory whenever a failure to meet the criteria of the other methods is encountered.

To insure that the product shipped continually meets the intent of the distribution requirements, a provision is made for packaging the output in groups of 5 units. This in effect furnishes the user either with random sample groups of 5 from a process which has been shown to be in satisfactory control (control chart and batch methods), or with specially selected groups of 5, the units in each of which have been chosen to meet a particular distribution pattern (three-cell method).

The following sections give the general character of the statistical models that have been set up for the three methods.

3.1 CONTROL CHART METHOD

The control chart method is intended for application where production comprises a reasonably steady succession of individual units or small groups of units from a common source; so that units or groups, as produced, may be kept in the order of their production and control chart techniques applied to test results. Under this method control charts are maintained for averages and ranges of samples of 5.

At the outset it is necessary to demonstrate an adequate degree of control of the product in order to be considered eligible for application of this method. Once eligibility has been established, a second and somewhat more lenient set of conditions is used to judge whether this eligibility is maintained. For convenience of reference these two sets of conditions are designated (a) Criterion I, for establishing eligibility (or for reestablishing it), and (b) Criterion II, for maintaining eligibility.

The control charts used in this section are an adaptation of the well-known Shewhart control charts for sample averages, $\bar{X}$, and sample ranges, R , for "control with respect to a given standard." Two modifications of the techniques customarily used in applying the control chart for $\bar{X}$ have been introduced: (a) a central band rather than a central line has been provided, and (b) a non-parametric requirement has been imposed on seven successive sample averages to limit the excursions of the product average from the nominal value. These modifications are related to the two acceptable distributions referred to in Fig. 2 and reproduced as dotted lines in Fig. 3.

The PA limits of Fig. 3 are the desired minimum and maximum values of the process average, and are the averages of the two acceptable distributions. As indicated in Fig. 2, the PA limits for the process average give the boundaries of the band, $N \pm \frac{1}{3}\sigma''$, where σ'' is the standard deviation of the two acceptable distributions.

To determine control limits of the control charts for $\bar{X}$ and R , consideration is given to the sampling distributions of averages of samples of 5 drawn from such acceptable distributions, as shown in Fig. 3, as well as to the sampling distribution of ranges of samples of 5 from these distributions.

The A5 limits of Fig. 3 (limits to be met by the average of a sample of 5) are 3-sigma control limits for averages of 5, given by

$$N \pm \left(\frac{1}{3}\sigma'' + 3 \frac{\sigma''}{\sqrt{5}} \right).$$

The R5 limit (limit to be met by the range of a sample of 5) is the

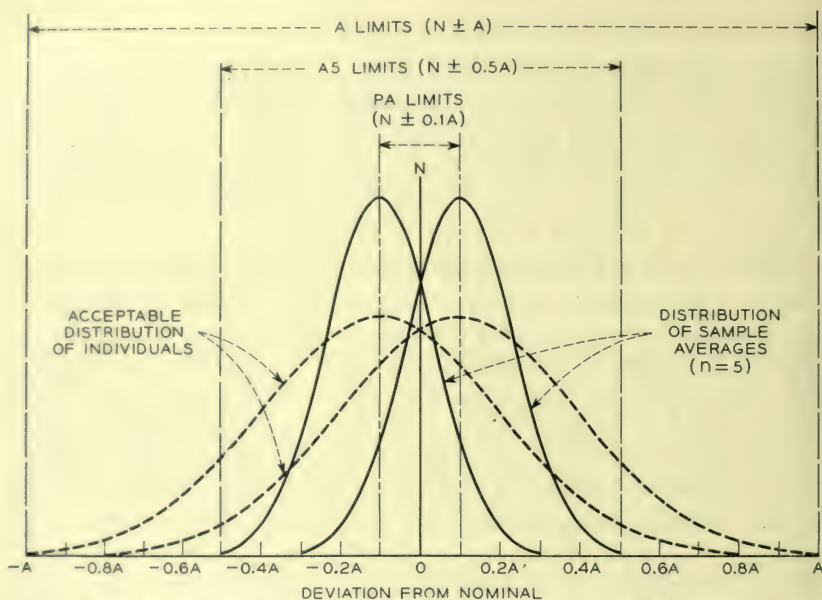


Fig. 3 — Basis of A5 limits and PA limits for sample averages.

customary upper 3-sigma control limit for ranges, $D_2\sigma''$ (where $D_2 = 4.918$ for samples of 5). Since the relation between the specified limits for individual units and σ'' is expressed by $A = \frac{1}{3}\sigma''$, the above limits are related to the specification limits as follows:

$$\text{PA limits} = N \pm 0.1A$$

$$\text{A5 limits} = N \pm 0.5A$$

$$\text{R5 limits} = 1.48A, \text{ max.}$$

At the outset and whenever eligibility to use the control chart method is lost, a 100 per cent inspection rate is required and acceptance is based on the three-cell method (discussed later). Units in groups of 5 are tested as successive samples, and subjected to the following criterion:

Criterion I — Establishing Eligibility.

Eligibility for use of the control chart method is established as soon as 7 consecutive samples or groups of 5 satisfy all the following conditions:

- (a) The averages all meet the A5 limits; and
- (b) The ranges all meet the R5 limit; and

- (c) The seven consecutive averages are not all outside the same PA limit (not all above the upper PA limit or all below the lower PA limit).

When eligibility to use the control chart method is established and so long as it is maintained, sampling inspection may be used. For this inspection, periodic samples of 5 are selected in accordance with a schedule which normally calls for measurement of about 10 per cent of the units produced. The lots represented by such samples are considered as conforming to the intent of the distribution requirements, and hence acceptable for this feature. Provisions are made for further reducing the inspection rate when consistently good control performance is evidenced. During the period of sampling, the following criterion applies to the sampling results:

Criterion II — Maintaining Eligibility

Eligibility for use of the control chart method is maintained so long as the results of the current sample satisfy all of the following conditions:

(a) The average either (1) meets the A5 limits; or (2) fails to meet the A5 limits but at the same time all of the 6 preceding consecutive averages meet the A5 limits; and

(b) The range either (1) meets the R5 limit; or (2) fails to meet the R5 limit but at the same time all of the 6 preceding consecutive ranges meet the R5 limit; and

(c) Seven consecutive averages (for the current sample and the 6 preceding samples) do not all fall outside the same PA limit.

(d) No major change is made in raw material, machine set-up or personnel, which may have a significant effect on the quality of the product.

Fig. 4 gives a control chart for averages and ranges of samples of 5 such as might be obtained under the control chart method. The first 20 points (Series A) indicate what might be expected in a series of samples if the process were controlled with its average, $\bar{X}'$, at N and its standard deviation, σ' , equal to the standard value, $0.3A$. The next 20 points (Series B) indicate the expected pattern of points if the process average suddenly shifted to a level about $0.25A$ above the nominal, while σ' remained unchanged. Both sets of points represent the result of random sampling experiments simulating the conditions stated. During Series A the first seven points meet Criterion I and would have established eligibility for using the control chart method. Eligibility would have been maintained for the balance of Series A. Starting with Series

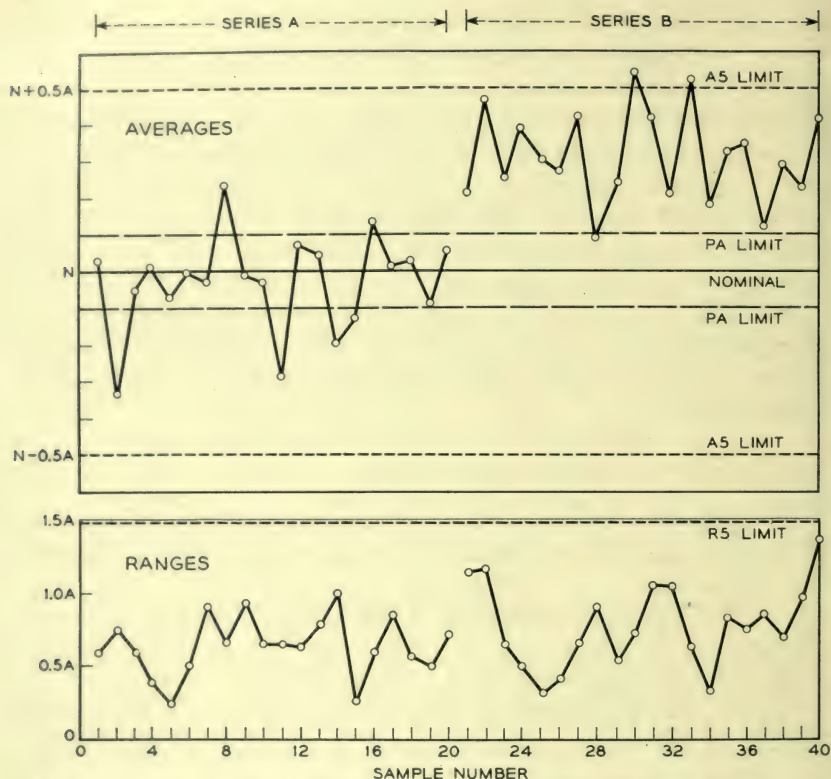


Fig. 4 — Control charts for averages and ranges, samples of 5 from a normal universe.

B, however, a failure to meet Criterion II would have been encountered on the seventh plotted point, which would have caused loss of eligibility to continue using the control chart method. This would have required a reversion to 100 per cent inspection and acceptance by the three-cell method until such time as Criterion I was again met.

Individual units in a lot accepted by the control chart method are to be selected at random from the lot in groups of 5 each, and the five units of each group are to be placed in a single package or otherwise associated so as to remain physically together until delivered to the user.

3.2 BATCH METHOD

For some component elements, units are produced intermittently in relatively large groups or batches and subjected collectively to the same manufacturing processes. Under these conditions individual units

or small groups of units do not follow one another through the process in time sequence. Here "order of production" has no meaning for individual units — the only order is the sequence of successive batches.

For this situation a batch method is provided whereby a substantial sample, normally 50 units, is selected at random from each batch. The average, $\bar{X}$, and the standard deviation, σ , of the sample are computed and the satisfactoriness of the distribution of the batch is determined by comparing (1) the sample average, and (2) the sample standard deviation with certain limits of allowable variation which have been established. As in the case of the control chart method, consideration is given to the results to be expected in samples from the two acceptable distributions already defined.

The A50 limits, the limits to be met by a sample average, are 3-sigma control limits for averages on either side of the process average band given by $N \pm (\frac{1}{3}\sigma'' + 3\sqrt{\sigma''}/\sqrt{50})$, which after substituting $\sigma'' = 0.3A$, gives

$$\text{A50 limits} = N \pm 0.23A.$$

The limit to be met by a sample standard deviation is the upper 3-sigma limit of a sampling distribution of standard deviations for samples of 50 from a universe having a standard deviation, σ'' . However, in practice the standard deviation is difficult to calculate. In order to simplify calculations, an estimated standard deviation is used instead, computed as follows: Divide the 50 values into random subgroups of 5; find the range, R , of each of the subgroups; compute the average range, $\bar{R}$, and multiply by 0.43. Thus σ (estimated) = $0.43\bar{R}$. Considering the sampling distribution of this statistic (0.43 times the average range for 10 samples of 5) as a substitute for σ , we make use of known theoretical relations between the distribution of $\bar{R}$ for samples of 5 and the standard deviation of the sampled universe. The upper 3-sigma limit of the sampling distribution of $\bar{R}$ (for 10 samples of 5) is given* by

$$2.326\sigma'' + 3 \left(\frac{0.864\sigma''}{\sqrt{10}} \right)$$

which, after substituting $\sigma'' = 0.3A$, gives 0.95A. From this the limit to be met by the sample standard deviation (estimated as $0.43\bar{R}$) is

$$\text{S50 limit} = 0.41A, \text{ max.}$$

The limits for the sample average must be met by each batch, but

A.S.T.M. Manual on Quality Control of Materials, Am. Soc. for Test. Mat., Phila., 1951; see related formula D₂, p. 114, and factors d₂ and d₃, p. 115.

provision is made for an occasional failure to meet the limit for the sample standard deviation. Thus the batch represented by a sample will be considered conforming to the intent of the specified distribution requirements, and hence acceptable for this feature if the sample meet the following criterion:

Criterion III — Batch Acceptability.

- (a) The average, $\bar{X}$, meets the A50 limits; and
- (b) The standard deviation, σ , either (1) meets the S50 limit, or (2) fails to meet the S50 limit but at the same time all of the six preceding consecutive standard deviations meet the S50 limit.

If a batch fails to meet this criterion, the batch must be inspected 100 per cent and acceptance is based on the three-cell method.

Packaging is handled in the same manner as for the control chart method; individual units in a batch accepted by the batch method are to be selected at random from the batch in groups of 5 each and packaged as such for delivery to the user.

3.3 THREE-CELL METHOD

Even with the best of conditions things may go wrong from time to time due to any one of a number of causes — changes in raw materials, irregularities in manual operations, faulty performance of processing equipment, etc. As a result, samples taken under either the control chart method or the batch method will sometimes fail to meet the criteria of the methods. In such times of trouble one solution would be to stop production, find the assignable cause, rectify it, and then resume manufacture. This solution, though perhaps ideal in one sense, may not be practical for several reasons:

- (a) Considerable time may elapse before the assignable cause is found and corrected;
- (b) Manufacturing schedules may be disrupted; and
- (c) No answer is provided to the question of what to do with the uncontrolled product already made.

What is needed, therefore, is a procedure for dealing with the finished units when the process is in trouble distribution-wise, a procedure which will permit shipment of some of the product and at the same time assure that the portions shipped will have a proper distribution.

To this end a selection procedure referred to as the three-cell method has been provided. Under this method each unit of product is measured for the characteristic in question and the conforming units are classified

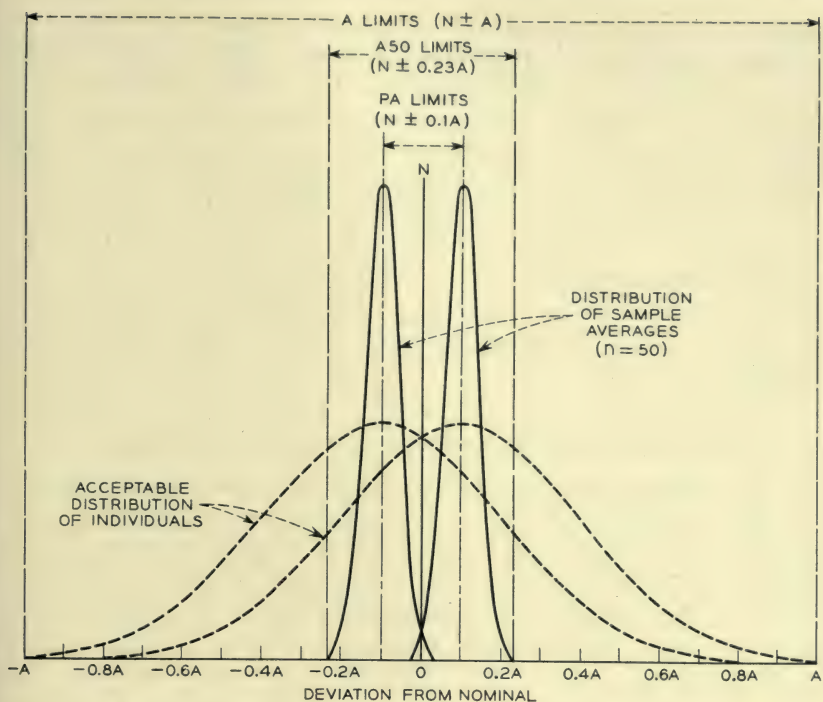


Fig. 5 — Basis of A50 limits for sample averages, batch method.

and sorted into three cells: lower, middle and upper, as shown in Fig. 6. Units are then selected from the three cells in groups of 5 and packaged, the five units in each package to be distributed among the three cells in accordance with one of the two distributions shown at the top of Fig. 6. These selected groups of five are maintained as packages of 5 in merchandise stock and in deliveries to the user.

Units in any cell which are in excess supply at any given time during production and which, therefore, cannot be packaged may be included with subsequent production provided each package of five satisfies one of the two distributions shown in Fig. 6.

For the three-cell method several matters were open to choice — the relative width of the three cells, the number of units in a package, and the required distribution of units in a package. Sampling experiments were run with groups of 4, 5, 6 and 12 and the net effects studied probability-wise for various possible kinds of quality situations that might be encountered in production.

With appropriate choices of these items it is possible to provide as-

surance that the units furnished from various segments of production will approximate the deviation pattern to be expected in random samples from an appropriately controlled process. The fundamental objective is to provide a parade of product-segments under the three-cell method that will continue to have distributions that meet the basic intent of the distribution requirements as deliveries or replacements are made. At the same time, it is desirable to make the three-cell method moderately more restrictive than the control chart method and the batch method in order to provide an incentive to attain a degree of control during production that will permit sampling. The choice of three equal cells, packages of 5, and the package distributions indicated in Fig. 6 were selected with these things in mind.

4.0 OPERATING CHARACTERISTICS OF CONTROL PROCEDURES

In the preceding section a description was given of the control procedures that are associated with specified distribution requirements. The most important question to be answered is: "What are the operating characteristics of these procedures?" In other words, how well will these procedures discriminate between quality that is good or bad, distribution wise? How effectively will they assure realization of the objectives for which they were set up?

For the sake of simplicity the statistical models used to evaluate the expected performance of the control methods are limited to Normal dis-

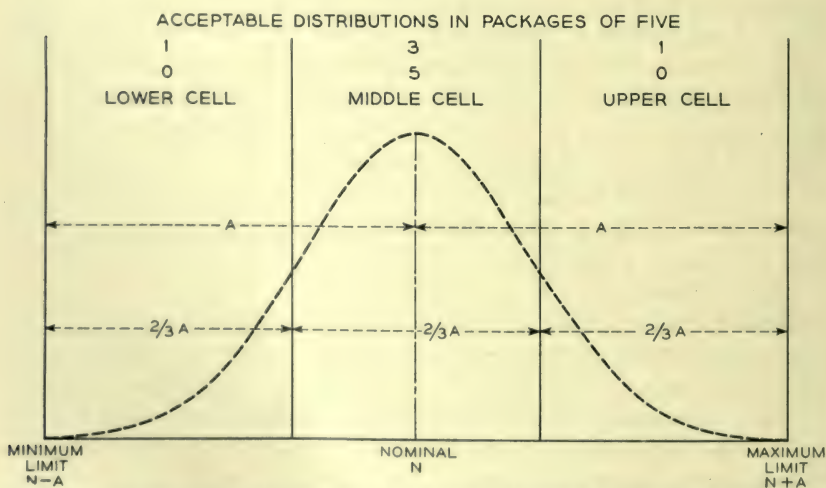


Fig. 6 — Acceptable distributions of units in packages of 5, three-cell method.

tributions.* Since a primary objective is to limit the displacement of the process average, $\bar{X}'$, from the nominal value, N , this parameter will be used as the independent variable against which the performance of the methods is computed. The curves to be shown will be referred to as "operating characteristic curves" or "OC curves" for the respective criteria of these methods, a term† that is applied generally to the related "probability of acceptance" curves for sampling inspection plans.

4.1 OPERATING CHARACTERISTICS OF THE CONTROL CHART METHOD

First, let us assume that we have a process the output of which has a Normal distribution with a standard deviation, σ' , equal to $0.3A$, which is the same as the standard value, σ'' , used in setting the specified limits. For this case the limits $N \pm A$ are $N \pm 3\frac{1}{3}\sigma'$. Ideally, the process average, $\bar{X}'$, should be centered at N , but by design it is agreed that a displacement of $\bar{X}'$ by an amount $0.1A$ from N should be acceptable. Conceivably the displacement could well be considerably larger than this, and the question arises as to how the two criteria of the control chart method will function for different magnitudes of displacement. With the model assumed, it is possible to work out an analytic answer to this question. For example, for any given displacement of $\bar{X}'$, from N , the resulting formula for P_I , the probability of meeting Criterion I, is:

$$P_I = \left[(1 - P_2^+ - P_2^-)^7 - (P_1^+)^7 - (P_1^-)^7 \right] \left[(1 - P_4)^7 \right] \quad (1)$$

and the formula‡ for P_{II} , the probability of meeting Criterion II is

$$\begin{aligned} P_{II} = & \left[P_0 + P_1^+ \left\{ 1 - (P_2^+ + P_1^-)^6 \right\} \right. \\ & + P_1^- \left\{ 1 - (P_2^- + P_1^-)^6 \right\} + P_2^+ \left\{ (1 - P_2^+ - P_2^-)^6 - (P_1^+)^6 \right\} \\ & \left. + P_2^- \left\{ (1 - P_2^+ - P_2^-)^6 - (P_1^-)^6 \right\} \right] \left[1 - P_4 + P_4(1 - P_4)^6 \right] \quad (2) \end{aligned}$$

* For other distributions this limitation is unimportant for those portions of the criteria that relate to sample averages since averages of samples from a non-Normal universe may ordinarily be considered to be distributed Normally. See W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Co., New York, 1931, pp. 180-184. But due among other things to lack of independence of $\bar{X}$ and R for skew distributions, the results given here should be considered only as reasonably close approximations for the degrees of non-Normality that may be encountered in practice.

† A term first used in the late 1930's by Capt. H. H. Zornig, of the Ballistic Research Laboratories, at Aberdeen Proving Ground.

‡ P_{II} is an unconditional probability in the sense that it does not involve the condition that previously there was a sequence of 7 samples satisfying Criterion I and no intervening sequence of 7 samples not satisfying Criterion II. P_{II} has been used as an approximation to, although possibly somewhat less than, the corresponding conditional probability. A similar consideration applies to P_I .

where each of the above P symbols designates the probability that a sample average (average of a random sample of 5 units) will fall in the band associated with that symbol in the following tabulation:

Symbol	Band
P_2^+	Above $+0.5A$
P_2^+	$+0.5A$ to $+0.1A$
P_0	$+0.1A$ to $-0.1A$
P_2^-	$-0.1A$ to $-0.5A$
P_2^-	Below $-0.5A$

and P_4 = probability that a sample range (range of a random sample of 5 units) will fall above its upper control limit, $1.48A$. ($P_4 = 0.0044$ for $\sigma' = 0.3A$.)

The OC curves for Criterion I and Criterion II computed from these formulas are given in Fig. 7(a), when $\sigma' = 0.3A$. These show how the control chart method serves on a probability basis as a band-pass filter for a manufacturing process, permitting the introduction of and allowing the continuance of sampling so long as the process average is maintained reasonably close to the nominal, and imposing an increasingly higher barrier to acceptance by sampling when the displacement of the process average from N is increased, thus forcing the use of the three-cell method. Examination of these curves shows that Criterion I is more stringent than Criterion II, as it should be.

Suppose for the moment that both Criterion I and Criterion II omitted condition (c) relating to seven successive sample averages. The OC curves in this case are shown in Fig. 7(b) with the designation "I, II, less c." It is seen that this requirement relating to seven successive averages is most important. Without it the criteria, particularly Criterion II, would be very ineffective in controlling excursions of the process average.

A second question of importance is: "What happens if the specified dispersion limits (A limits) are improperly set or if the process dispersion changes to produce this effect?" In other words, what are the operating characteristics of the procedure when $\sigma' \neq 0.3A$? This is shown in Figs. 7(c) and 7(d) for Criterion I and Criterion II respectively, where the values of process standard deviation (σ') are expressed as fractional values of A .

4.2 OPERATING CHARACTERISTICS OF THE BATCH METHOD

For the batch method the procedure is essentially a lot acceptance procedure and except for permitting an occasional failure of the sample

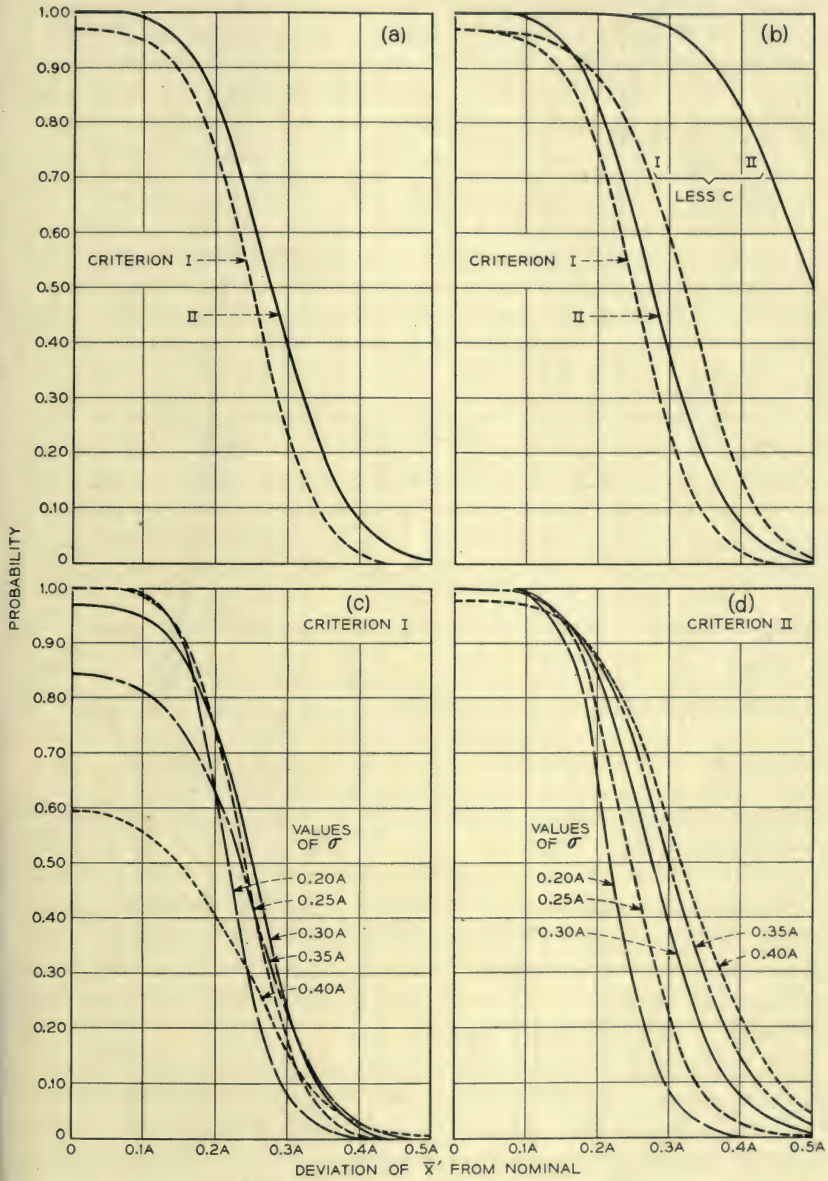


Fig. 7 — Operating characteristic curves for control chart method.

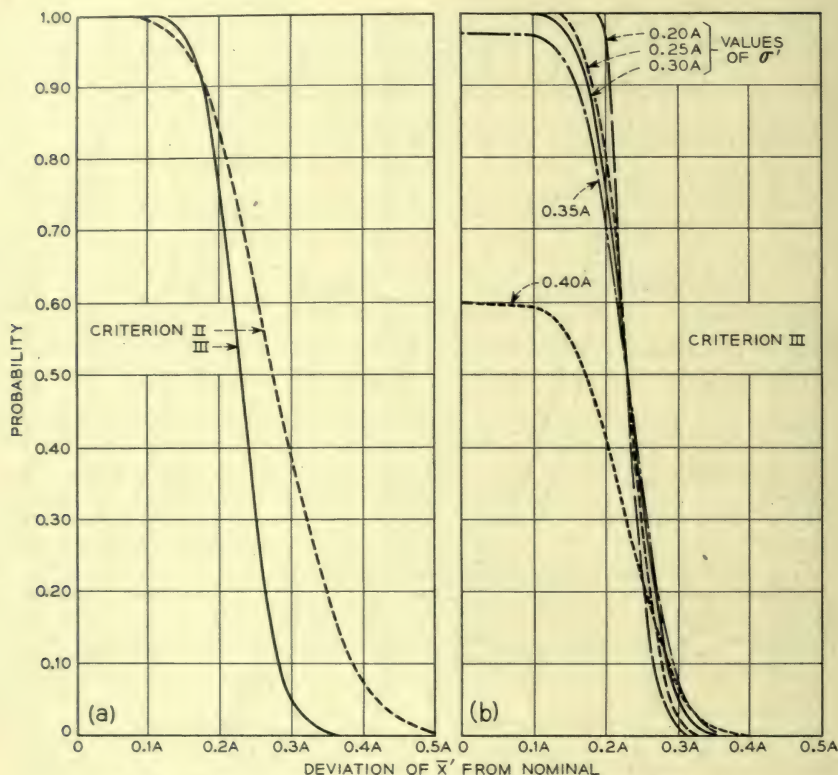


Fig. 8 — Operating characteristic curves for batch method.

standard deviation to meet its limit, each lot is judged solely on the data obtained from the sample from the lot. The OC curve for the batch method (Criterion III) is shown in Fig. 8(a) for the case where $\sigma' = 0.3A$. For comparison purposes the corresponding curve for Criterion II of the control chart method is also shown. It is noted that the batch method gives a somewhat sharper discrimination between good and bad distributions than Criterion II of the control chart method, due primarily to the use of a relatively larger sample.

Fig. 8(b) shows how the OC curve is modified for other values of process standard deviation, σ' . It is seen that the batch method is relatively less sensitive than the control chart method to changes in σ' provided the deviation from a standard value of $\sigma'' = 0.3A$ is not too great.

4.3 CHARACTERISTICS OF THE THREE-CELL METHOD

The operating characteristics of the three-cell method cannot be evaluated probability-wise in the manner given for the other two methods.

However, the manner in which the three-cell method serves as a continuous corrective influence over the distribution of delivered product can be indicated by a few diagrams, for all of which a Normal distribution with $\sigma' = 0.3A$ is assumed.

The running average of small segments of product delivered in packages of 5 is held closer to the nominal by the three-cell method than by the control chart method or the batch method, even when the process is statistically controlled at the nominal. A comparison with the control chart method is illustrated in Fig. 9. In the upper chart are shown averages of random samples of 5 units each, plotted on a control chart with A5 and PA limits. These are samples obtained experimentally from a Normal distribution whose average, $\bar{X}'$, was at the nominal for the first 20 samples (Series A). For the next 20 samples (Series B) the average was 0.15A above the nominal and for the last 20 samples (Series C) the average was 0.15A below the nominal. The same units were then classified and packaged by the three-cell method. In this experiment, as the units of each sample were classified, as many units were packaged in 1-3-1 or 0-5-0 distributions as possible. Of the first 100 units (20 groups of 5), 95 were packaged. After 200 units (40 groups of 5) were sorted into cells,

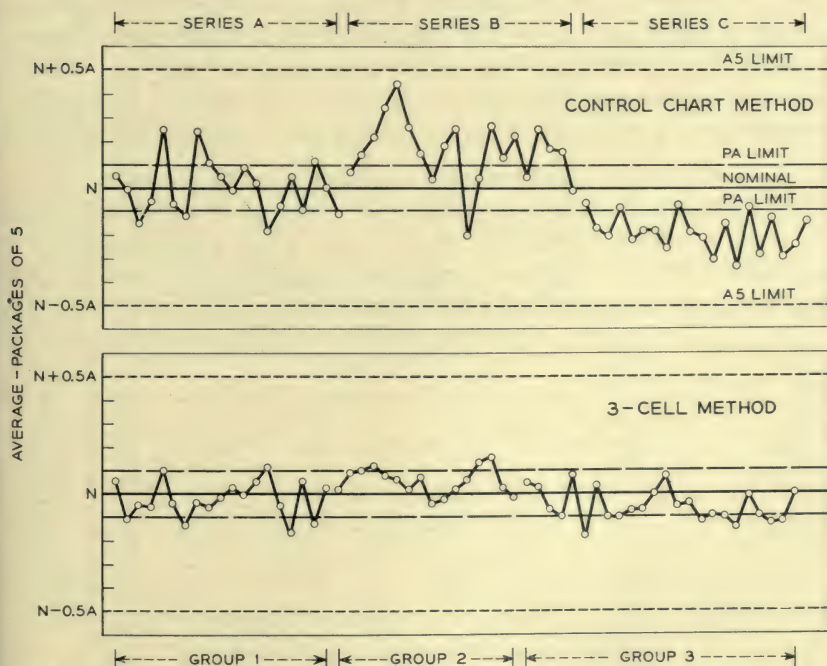


Fig. 9 — Comparison of control chart and three-cell method, averages of packages of 5.

175 were packaged. Of the 25 units not packaged, 24 were in the upper cell and 1 unit did not meet the A limits. At the end of the experiment 295 of the 300 units had been packaged. Of the 5 units not packaged, 3 remained in the upper cell, and 2 did not meet the A limits. The averages of the packages are shown in the lower chart of Fig. 9. For comparison, the PA limits are also shown on this chart. It is apparent from these two charts that the three-cell method yields packages whose averages are held closer to the nominal than are averages for packages from the control chart method.

The corrective effect of the three-cell method is further illustrated by Fig. 10, which shows the average of product packaged by the three-cell method as a function of the process average. This curve is for "long term" conditions, that is, it represents the expectancy for any given level of process average. This corrective effect is purchased at the expense of not packaging a portion of the product while the process average is not at the nominal value. However, as already noted, the unpackaged portion may be packaged with subsequent product if the process average subsequently deviates from the nominal in the opposite direction.

The percentage that can be packaged is also shown in Fig. 10 as a function of the deviation of the process average from the nominal. It should be noted that this curve also represents the expectancy for any given level of process average. Of course, continued production at a fixed level other than nominal would result in a steadily growing accumulation of unpackaged units, a situation that would call for corrective

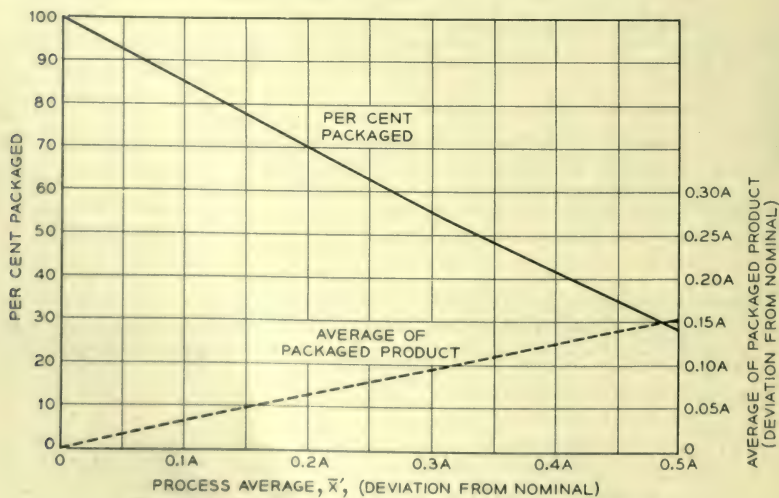


Fig. 10 — Expectancy curves for three-cell method.

action on the process. Close to 100 per cent packaging can be assured by introducing a negative bias in the process to compensate for the effect of a prior positive bias, and vice versa.

5.0 CONCLUSIONS

The L3 system's need for holding transmission performance close to the design center, both within short segments and over the full span of the transcontinental line, has called for a high degree of statistical uniformity of critical characteristics of component elements. The statistical quality control methods are imposed from the point of view of the user in the interests of the over-all economy of system design. The control procedures are designed to provide at all times a parade of suitably distributed batches of production units, and at the same time to furnish incentives for controlling manufacturing processes at the design center.

Any enterprise of this kind, involves the closest of interplay and adjustment between design and production interests. Many cases of incompatibility of design desires and production capabilities had to be cleared in the early stages of the work. Intensive process quality control work and the development of a number of ingenious processing techniques on the part of the Western Electric Company have contributed greatly to what has been achieved. Experience will undoubtedly indicate the need for some refinements or adjustments in the plan.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to members of the Western Electric Company's engineering organization for cooperation in the development of the general plan, to Miss M. N. Torrey and Dr. R. B. Murphy for development work on statistical features of alternate and final plans, and to the Misses E. F. Lockey and J. Zagrodnick for conducting sampling experiments and making computations.

The L3 Coaxial System

Application of Quality Control Requirements in the Manufacture of Components

By R. F. GARRETT, T. L. TUFFNELL and R. A. WADDELL

(Manuscript received April 27, 1953)

The application of quality control procedures, in addition to conventional maximum and minimum limits, is an important factor in the manufacture of components for the L3 carrier repeaters. In this application, control chart techniques are used for providing assurance that the average of each characteristic subject to control is held close to a desired value and that, collectively, individual units have a desired distribution about this average value. The three-cell method is frequently used under certain conditions encountered in the manufacture of these components when sampling procedures cannot be applied. This method consists of measuring each unit of product, classifying conforming units into one of three cells and the selection of groups of five units each to provide the desired distribution. Case histories of a number of factory applications of these methods are presented.

1.0 INTRODUCTION

1.1 GENERAL

Statistical quality control methods are well known and useful industrial tools for economically controlling quality during manufacture. These methods have as one of their goals the shipment of product meeting the end requirements for a particular quality characteristic. The application of such methods also makes possible the delivery of a product whose quality is statistically uniform as, for example, having a distribution whose average is maintained consistently close to the design center. Bell Telephone Laboratories engineers have made use of these principles in the development of the new L3 long distance carrier system which will employ hundreds of repeaters in tandem.

An important factor in the application of statistical quality control methods is the specification of distribution requirements in addition to

conventional maximum and minimum limits for the characteristics of many of the components manufactured for the new repeaters. The introduction of distribution criteria where maximum and minimum are customarily specified requires a number of operations which are supplementary to normal procedures. The relatively simple act of identifying good product becomes complicated by the need of more extensive measurements, the recording of data, computations, plotting of charts and the active participation of technical personnel in the administration of the procedures.

The first step in the program was the development of practical statistical quality control techniques which would be applied under the special circumstances attending the design and manufacture of components of L3 carrier repeaters. This required careful study by Bell Telephone Laboratories and the Western Electric Company and resulted in the development of a general specification which provides procedures and criteria for maintaining the average value of a quality characteristic close to a nominal value and for obtaining as nearly as possible a random distribution of individual values around the nominal. The purpose of this paper is to discuss the procedures thus developed with emphasis on their relationship to manufacturing processes and to describe the problems encountered and solved in the course of application to the manufacture of components of L3 carrier repeaters. Detailed mathematical derivations and terms will be generally omitted since the theories underlying the principles involved are covered by another article.¹

1.2 PARTICIPATION IN DEVELOPMENT

Normal practice in the creation of new product designs is for the development and construction of the first models to be handled by the design engineers. This work usually includes discussions with the manufacturing organization in order to minimize the costs and to utilize existing or most effective manufacturing facilities and various preferred or stocked materials. In the case of the critical components of the L3 carrier amplifiers, a design change in one component resulting from the transition from development to production requires especially close study and may require an adjustment in other components in order to compensate for the one being changed. Knowledge of the behavior of regular manufacturing facilities and methods used in the fabrication of preproduction units provides considerable assistance in establishing specification limits which are compatible with the product design and manufacturing process capabilities.

1.3 SEQUENCE OF MANUFACTURING OPERATIONS

The application of distribution requirements places added emphasis on the proper sequence of the various operations required for the fabrication of the product. Normally, any assembly or finishing operation following a process adjustment of a particular characteristic is designed to keep that characteristic within maximum and minimum limits in the final state. Such procedures often fail to satisfy the desired distribution and it is necessary to rearrange the sequence of operations. Once the proper sequence is established it must be rigidly maintained.

1.4 TESTING

As a result of refinements in design and in production methods employed for the critical components of the L3 amplifiers, the design engineer has in many instances been able to specify limits closer than ever before attained. The specification of such close limits may tax the precision of factory testing equipment and in many cases it has been necessary to develop and construct new electrical and mechanical inspection facilities. Measurement reproducibility as well as accuracy in terms of absolute values is important since the measuring instrument ordinarily indicates variations in repetitive readings, even though the product being measured remains constant. Once the characteristics of the measuring instrument are determined and used as a basis for the specification of limits, the measuring facility becomes an important part of the distribution control system and must be controlled the same as all other elements of the system. This means careful watch over the maintenance of factory inspection facilities so that these characteristics are controlled. Obviously, an adjustment made on the measuring instrument in the course of regular maintenance which introduces a significant bias or shift, even though well within accuracy limits, may have to be taken into consideration in the use of the instrument. One method of minimizing this problem is to employ stable fixed standards whose characteristics are numerically equal to or near the nominal value of the products being tested. Such standards can be used for either calibrating the instrument or as comparison standards in the actual measurement of the product. These auxiliary standards must still be periodically checked and extreme care taken to prevent any shift in their characteristics.

2.0 QUALITY CONTROL METHODS

2.1 CONVENTIONAL STATISTICAL QUALITY CONTROL METHODS

The concept of control as used here includes the use of data resulting from measurements made on product produced under the same essential

conditions. The curve which can be used to represent the observed frequency distribution of data obtained under controlled conditions may, for most practical purposes, be illustrated by the shape shown in Fig. 1. The characteristics of this curve are mathematically represented by the average $\bar{X}$ (arithmetic mean) and the standard deviation σ (the root-mean-square deviation of individual values from their average $\bar{X}$).

The control chart techniques as originally developed by Walter A. Shewhart² of Bell Telephone Laboratories provide economical methods for measuring and evaluating the characteristics of such distributions. In practice they are useful in obtaining an estimate of the capabilities of manufacturing processes, sometimes referred to as the "natural tolerances" of the process and in maintaining control at that quality level. In general, a process having a controlled distribution with an average $\bar{X}$

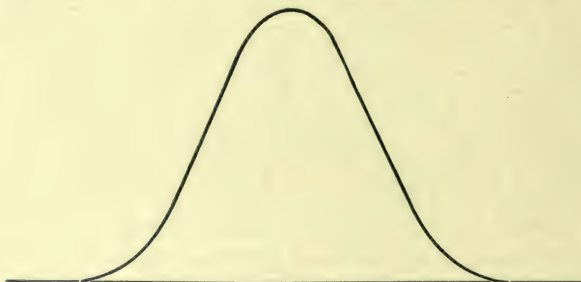


Fig. 1 — The ideal frequency distribution for observations obtained under controlled conditions.

and a standard deviation σ will result in practically all of the individual units of product falling within the band $\bar{X} \pm 3\sigma$.

Conventional quality control techniques are used by both the design and manufacturing engineers. The design engineer, in order to specify tolerances compatible with the design needs and the capabilities of economical commercial manufacture, applies the techniques to a reasonable number of preproduction models or possibly to a limited quantity of initial regular production. The manufacturing engineer in turn uses the techniques for determining the capabilities of existing or new manufacturing facilities in order to select the most effective facilities and methods. The techniques are also effective tools for locating and eliminating assignable causes of manufacturing variations, while their continued use as a regular part of the manufacturing process provides an excellent contribution to effective quality control. In these applications there usually exists a substantial margin between the $\pm 3\sigma$ variation around the nominal value and the specification limits. This is illustrated

as specification limits, 1-1 in Fig. 2. The margin is attained either by improvement of the manufacturing processes or by widening the specification limits. This means that occasional out-of-control conditions can be indicated by the control chart without the factory being faced with a shut down in production, provided that corrective measures are taken before the magnitude of the deviations results in any significant quantity of the product failing to meet the specification limits.

The contrasting situation occurs where the spread between the maximum and minimum specification limits becomes equal to or less than the 6-sigma spread of the manufacturing process, as shown in Fig. 2. For specification limits 2-2, production must be stopped every time an out-of-control condition is indicated or 100 per cent inspection introduced until the cause of the condition is located and eliminated. In the case of specification limits 3-3 in Fig. 2, 100 per cent inspection must be continuously applied in order to eliminate non-conforming product as shown in the shaded portion under the distribution curve.

2.2 SPECIAL STATISTICAL QUALITY CONTROL METHODS

The special statistical quality control techniques developed for L3 carrier components present a special problem. Here they become an actual part of the product specification, rather than an aid in meeting it, although conventional methods may still be useful for process control. It was recognized that any attempt to express specification limits in terms of a $\pm 3\sigma$ variation around the nominal would have to include

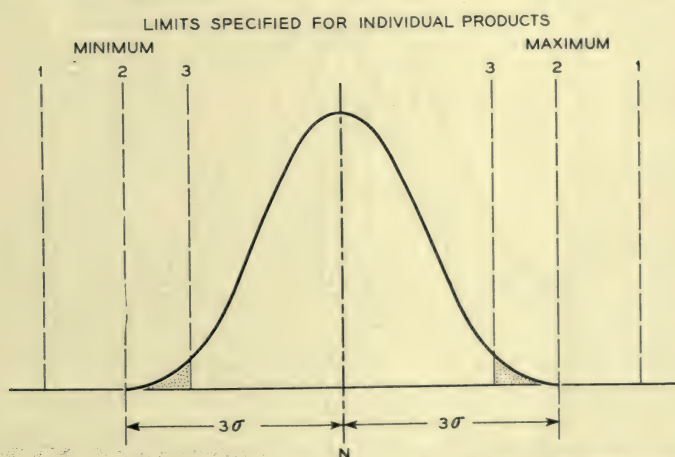


Fig. 2 — The relations between a frequency distribution of individual units of product and various specified maximum and minimum limits.

so that large inventories will not accumulate in the factory and delivery commitments can be met. Such a method requires the measurement and classification of the product into groups or cells, followed by the selection of units from such cells in accordance with a required distribution. Although this method requires 100 per cent inspection instead of sampling inspection permitted by the use of control charts it was estimated that this method would be used extensively until new manufacturing processes were thoroughly proven and sufficient data collected to fully establish a process capability.

2.3 THE GENERAL SPECIFICATION

Anticipated application required consideration and development of methods for the wide variety of manufacturing conditions likely to be encountered in the production of such items as coils, condensers, resistors and vacuum tubes. These products may be manufactured at widely separated factories for assembly in equipment at still another location. In order to assist in the description of the factory applications, to be presented later, the methods developed which are referred to as "distribution requirements" in the general specifications are listed and briefly described below. With the exception of the Records method, distribution requirements are applied to only one characteristic of a product. If application to more than one characteristic is desirable a method suitable for use as a basis for shipment of the product is selected for the most important characteristic and the Records method is specified for the remaining characteristics.

1. Continuous production method.
2. Batch production method.
3. Three-cell method.
4. Records method.

2.31 *Continuous Production Method*

This method is for application where production comprises a reasonably steady succession of individual units or small groups of units of product from a common source so that individual units as produced may be kept in the order of their production. The criteria of this method apply to a series of units or groups of units of product arranged in the order of their production and are based on the use of control charts for averages and ranges for samples of 5 units each. Examples of typical control charts are shown in Figs. 4 and 5.

The inclusion of an allowance for the long time variation in the level

the various control limits for factory use is a relatively simple operation. A typical example is the 1507A Inductor where the inductance requirement is expressed as 283.9 ± 20 microhenries.

Then the "C" limits = $283.9 \pm 0.5 \times 20$, = 293.9 max., 273.9 min. microhenries, and the "E" limit = 1.48×20 = 29.6 microhenries.

Since the distribution requirements are a part of the specification, the product must meet rather exact criteria in terms of the limiting conditions previously described, even though they may appear complicated in terms of the usual manufacturing procedures. In the case of the Continuous production method the criteria are applied to the samples of 5 units each, in two steps, (1) to establish eligibility, either at the start of production or when eligibility is lost and (2) to maintain eligibility when once established. Prior to the establishment of eligibility or when eligibility is lost the three-cell method is applied.

Criterion 1

Eligibility is established as soon as 7 consecutive samples of 5 units satisfy the following:

- (a) The averages, $\bar{X}$, all fall within the *C* limits.
- (b) The ranges, *R*, all fall below the *E* limit.
- (c) Seven consecutive averages, $\bar{X}$, are not all outside the same *D* limit (not all above the upper *D* limit or all below the lower *D* limit.)

Criterion 2

Eligibility is maintained as long as each current sample of 5 units satisfies the following:

- (a) The average, $\bar{X}$, either
 1. falls within the *C* limits or
 2. falls outside the *C* limits but at the same time all of the 6 preceding consecutive averages fall within the *C* limits.
- (b) The range *R* either
 1. falls within the *E* limit or
 2. falls outside the *E* limit but at the same time all of the 6 preceding consecutive ranges fall within the *E* limit.
- (c) Seven consecutive averages $\bar{X}$ (the current sample and the six preceding samples) do not fall outside the same *D* limit.

2.32 Batch Production Method

This method is for application where production consists of intermittent batches of 50 or more units, all of which have been made under the

same essential conditions with respect to materials, parts, workmanship and processing. The criteria of this method apply to a series of batches or lots of product arranged in the order of their production and are based on the use of control charts for averages and standard deviations for samples of 50 units each.

The control limits for averages are for 50 units of product and also include an allowance for the long time variation in the level of the process average. They are represented by a band $N \pm 0.23A$ which is equal to the customary³ 3-sigma control limits for averages of 50 units placed outside a band corresponding to $N \pm 0.1A$ ($N \pm \frac{1}{3}\sigma$) and are designated in the general specification as "C" limits. In this method the "E" limit in the specification is the control limit for the standard deviation (σ) of a sample of 50 units and equals $0.41A$.

For this method a batch, represented by the sample, is considered conforming if the sample meets the following criterion.

- (a) The average $\bar{X}$, falls within the C limits.
- (b) The standard deviation, σ , either:
 1. falls within the E limit; or
 2. falls outside the E limit but at the same time all of the 6 preceding consecutive standard deviations fall within the E limit.

2.33 Three-Cell Method

This method is for application whenever the product fails to meet the criteria of the first two methods or where the manufacturing conditions dictate its use. When maximum and minimum limits are specified the method consists of testing or measuring each unit of product and classifying conforming units in one of three cells: Lower, Middle or Upper.

The distribution is shown graphically in Fig. 6. Groups of 5 units of product are selected from this classification so as to meet one of the two distributions given in Table I and maintained as such for shipment.

In order to provide a graphic record of whether or not the manufacturing process is in statistical control, this method includes procedures for obtaining control chart data based on samples of product taken from regular production before classification.

2.34 Records Method:

This method is for application when it is not practicable to impose the requirements of the other methods but it is still desirable to systematically maintain a graphic history of the measurements made on a quality characteristic. Action required to maintain the distribution of

the product within the limits implied by the control charts is left to the initiative of personnel administering the application of the method.

The records are based on measurements on samples of 5 units each selected at random prior to the regular inspection of individual units for conformance. The number of such samples should be adequate to provide a control chart record of the quality of the product.

In all of the methods described except the one requiring records only, the final product is packaged in groups of 5 units for shipment. When an order calls for a quantity other than an integral multiple of 5, the fractional part of the group shall consist of units randomly selected from a group which has been previously packaged.

3.0 FACTORY APPLICATION

3.1 GENERAL

The Distribution Requirements are being applied to 91 components consisting of 31 different codes of product used in the L3 carrier repeaters. In addition, Record methods are applied to 66 characteristics of the line and office amplifiers.

The components involved are identified in Table II.

These products require the maintenance of over 100 control charts in the factory in a form suitable for reproduction since engineering review of the results is necessary for adjustment of limits. Many of the charts are prepared and handled by factory personnel who required and received training in order to provide assurance of accurate results. Comprehensive instructions were prepared in order to assist in this training.

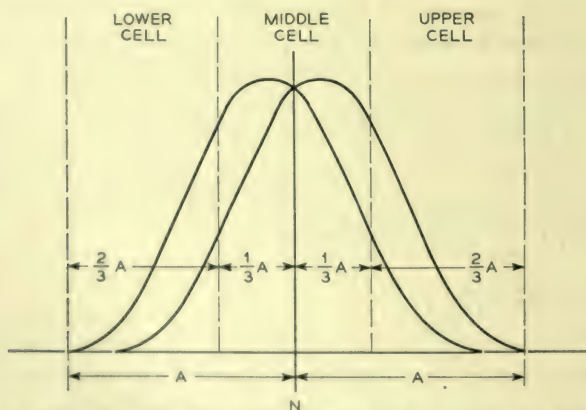


Fig. 6—A three-cell pattern superimposed on a frequency distribution of individual units of product when the average is allowed to vary $\pm \frac{1}{3}\sigma$.

TABLE I

Distribution	Number of Units		
	Lower Cell	Middle Cell	Upper Cell
1	1	3	1
2	0	5	0

A number of special forms were developed to facilitate the preparation of control charts, two of which are shown in Fig. 4 and Fig. 5. Since these control charts are directly related to the manufacture of the product, the cost of their preparation is included in the cost of the product, either by having the operation handled by production personnel or by the application of appropriate cost factors. This required in many cases the preparation of specific instructions to factory personnel in the form of manufacturing layouts which are also used as a basis for computing manufacturing costs and wage incentive rates.

One of the most serious problems in the application of these methods is the accumulation of large inventories of product when the manufacturing processes fail to meet the distribution requirements. In the case of regular maximum and minimum limits, procedures for the disposal of non-conforming product are direct and well known. In the case of distribution requirements however, there may be substantial quantities of product which meet the "A" limits but fail to meet the required distribution. Product which originally cannot be packaged must be put aside with the hope that product subsequently manufactured will result in combinations which will meet the required distribution. The discussion of applications to the products listed in Table II will show that a great measure of success has been realized in the solution of this problem.

TABLE II

Description	Number of Different Codes	Number of Different Repeater Components
Controlled pitch single layer coils (AWA, 302, 320, 1500, 1507 type).....	18	18
Mica capacitors (500 type).....	4	21
Quartz disc capacitors (505 type).....	3	3
Carbon deposited resistors (200 type).....	2	45
Quartz input-output transformer (2504 type).....	1	1
Vacuum tubes.....	3	3
Total.....	31	91

3.2 APPLICATION TO COILS

3.21 *Single Layer Type Inductors*

The problems encountered in the application of quality control techniques to the various codes of the single layer type coils follow a similar pattern so that a discussion of any one code covers many of the conditions common to the entire group. The early application of distribution requirements to L3 carrier coil characteristics was associated with the development of the 1500-type inductor commonly referred to as the "Splitting Coil". This coil derives its name from its primary circuit function of separating or splitting the transformer winding capacitance from the input capacitance of the amplifier vacuum tube. The circuit in which this coil is used permits an estimated variation in inductance from a desired value of approximately $\pm 1/4$ per cent, which includes manufacturing deviation, aging and temperature effects. A coil of this type having such close tolerances could not be produced economically with normal manufacturing methods. At this point the use of control chart methods played an important part in the development of tolerances and measuring techniques. The control chart techniques combined with the simultaneous development of manufacturing methods and sources of critical raw material have permitted the widening of limits of coils under the general distribution specification using "A" limits of ± 1 per cent in place of the originally estimated limits of $\pm 1/4$ per cent without a controlled distribution.

The splitting coil consists of a single layer winding wound on a ceramic form under a tension of 50 to 75 per cent of the wire breaking strength. The winding is terminated in lead wires which have been secured in holes located in the ceramic core transverse to the axis of the coil. The use of the ceramic core on this coil has reduced the temperature and aging effects to the point of becoming negligible, due to the nature of the core material and winding conditions. Major factors affecting the inductance variation are as follows:

a. Diameter of wire — standard commercial tolerances on wire of the sizes used, equalling ± 0.0003 " will produce a variation in coil inductance of approximately ± 0.15 per cent.

b. Diameter of core — a variation of core diameter of 0.0005" will produce a variation of approximately ± 0.5 per cent in the coil inductance.

c. Length of winding — a variation in the nominal length of winding of ± 0.001 " will result in a variation in inductance of about ± 0.13 per cent.

During the initial period of manufacture of these inductors, it was found that the seriousness of the variations in inductance, contributed by the wire diameter tolerance, could be minimized by selection of the wire stock. Experience has indicated that because of the small amount of wire used per coil, this procedure can be followed without undue losses.

The variations contributed by the core diameter tolerance have been minimized by the introduction of a ground core having a diameter tolerance of $\pm 0.0002''$. By virtue of the manufacturing methods introduced by the core manufacturer to meet this close diameter tolerance, a fairly normal distribution of the core diameter was obtained.

In addition to the variables due to the material used for these coils there were substantial variations in inductance inherent in the winding machines and operator winding techniques particularly in the attachment of the wire to the terminals. Early control charts showed an erratic



Fig. 7 — A shadowgraph of a coil winding showing irregularities in the wire spacing due to non-uniformity in pitch control of winding machine.

variation of inductance which was traced chiefly to backlash in the winding machine.

Fig. 7 is a shadowgraph of the coil winding showing the irregularities in the wire spacing due to non-uniformity in the pitch control of the winding machine. In addition to this variation, trouble was experienced in maintaining pitch control at the end of the winding due to the necessity of sanding the lead before splicing to the terminal. This latter variation was brought under control by the introduction of a type of wire which can be soldered to the terminal without mechanically removing the insulation, thus avoiding the loss of the winding machine accuracy of spacing and tension. After numerous machine modifications, satisfactory control of the winding pitch has been obtained so that with normal variations, the inductance of the coils is within the distribution requirements of the specification. This uniformity of winding is illustrated in the shadowgraph in Fig. 8, which is a typical example of current product. The type of machine modifications required to produce this improvement in product control included the following:

1. Eliminate slack in gear train by introduction of loaded gear drive.
2. Provide a constant force spring load on the worm drive so that the distributor exactly follows the cam motion.
3. Miniaturize wire guide assembly so that wire as placed on winding core will more closely follow the distributor motion.
4. Improve spindle bearings and distributor alignment.
5. Provide improved shut-off counter so that machine will give exact winding turns without overrun.

Although it might appear from the above discussion that winding coils on a controlled pitch basis will automatically result in the product having a normal distribution around a specified nominal inductance, the desired result can only be attained by continued vigilance, as many unpredictable variations will creep into the process from time to time. Gradual changes in control are readily observed from the charts, and in



Fig. 8 — A shadowgraph of a coil winding showing uniformity of winding necessary to meet distribution requirements for inductance.

many cases, are enough indication to show need of replacement or repair of worn winding-machine members which control the winding pitch of the coil.

The 320-type retardation coil used in the cosine equalizer of the L3 system, requires all of the control vigilance of the splitting coil and, in addition, has introduced the necessity for even further winding precautions. Although this coil has the same per cent tolerance as some of the other coils manufactured under this specification, it is more difficult to control, since it has only about one-tenth as many turns in the winding as most of the other coils. The major inductance variations in this coil, therefore, are controlled by the variation in the end turn location and the winder checks the inductance by a continuous sampling of the product during the winding procedure. It is thus possible to promptly make necessary modifications in the location of the final turn to maintain the inductance distribution within requirements. Experience has shown that normal variations from operator to operator can run as high as $\frac{1}{2}$ per cent on coils having a tolerance of 1 per cent, until the operator

becomes sufficiently skilled in the terminating methods. Thus, in order to obtain a product having normal distribution, close control of the winding operation must be maintained.

In addition to the above-mentioned single layer type coils, there is another general family of coils wherein distribution controls are useful during the initial period of development. The 1507 type, referred to as L-R inductors, are precise low-Q single-layer inductors wound with resistance wire on closely-controlled, low temperature-coefficient core tubes. These inductors are used in the input feedback network and as plate-feed inductors in the L3 amplifier, and are units which combine resistance and inductance in a single product. The stability of these coils is obtained by winding on a ceramic core form, similar to that used in the "splitting" coil, and using a low temperature-coefficient resistance wire in the winding. Simultaneous control of the resistance and inductance of this coil is obtained by the calibration of the winding machine setup to fit the resistivity of the wire used on the coil. Once the required turns and pitch have been determined for a given spool of wire, experience has shown that the resistance control will remain valid until the wire from the spool is exhausted.

The problem of making an accurate electrical measurement of a low Q inductor is a difficult one to resolve and it is here that an accurate control of testing methods by means of statistical analysis was employed. When an inductance (L_0) and a resistance (R) occur as series elements and the combination is shunted by the residual capacitance (C) across the bridge terminals, the resulting measured inductance is equal to that of the original inductance minus the product of the capacitance and the square of the resistance.

$$\text{Measured inductance} = L_0 - CR^2$$

The above equation is generally true in the case of coils where the inductive reactance is small compared to the resistance. If standard test procedures were followed to make the measurement of inductance, using the familiar Maxwell Bridges,⁴ an error would be introduced due to residual bridge and test jig capacitance. This would make the measured value considerably different from the coil inductance and would be dependent upon this residual capacitance and the coil resistance which could not be controlled with any degree of uniformity. Parallel resonating capacitance techniques are inadequate for the accuracies required, since the shunt equivalent of a very low-Q inductance is equal to that inductance divided by Q^2 . Determination of inductance by this method, while better than the Maxwell Bridge method, is subject to errors due

to the inadequate readability of the capacitance standard of available test facilities.

Since the theoretical values of resistance and inductance are specified by design, the following methods of determining actual design parameters are used:

A. Determine mathematically, or by experiment, the number of turns of a given resistance wire wound on the core tube to produce the required resistance.

B. Wind high Q inductors using copper wire of the same diameter as the resistance wire and having the same number of turns as determined in Step A, varying the winding pitch to produce the required inductance.

C. Wind low Q inductors with the wire and number of turns of Step A and the pitch as determined in Step B.

D. Make Maxwell Bridge inductance measurements for the two groups of low and high Q inductors. The two sets of readings are then subjected to a statistical analysis to insure that the processes are in control so that the data may be used in selecting a nominal coil from the low Q group to use as a reference standard. While it is true that the apparent value of the resistance-wound inductors will be appreciably less than those of the copper wire variety, due to the CR^2 correction, the nominal intrinsic inductances of the two groups of coils will be the same, due to their identical physical geometry.

Thus, by a method of comparison of two sets of readings, each of which is within process control, it is possible to select coils which can be used as reference standards. Any measurement which is made on a bridge set up from this reference standard will require correction of the inductance reading to compensate for the resistance variation. Self-compensating bridge methods have been developed which make it possible to read with good accuracy, directly on the bridge, the low Q coils covered by this family of inductors.

During the initial production of the 1507-type inductors, it was found expedient, for easier maintenance of control of the resistance and inductance variations, to use a nine-cell post-office method (three-cell by three-cell) which permits selection of product in any one of five ways using diametral axes, yet still maintaining the specification distribution covered by the three-cell method. The nine-cell method has gradually become less important, as experience has been gained with the use of the winding machines and control of winding operator variations.

Table III gives the over-all results of production of single-layer inductors and retardation coils, over a period of approximately one year. It can be seen that the design and production facilities are such that it is possible to manufacture a product within the requirements imposed by

TABLE III — YIELD OF COILS PRODUCED UNDER THE DISTRIBUTION REQUIREMENTS

Code	Nominal Values L—Microhenries R—Ohms	Tolerance Per cent ±	Total Production	Accumulated Un- packaged Product
AWA21 retard.....	L—191.61	1.0	514	90
AWA22 retard.....	L—166.2	2.0	342	6
302AW retard.....	L—1.58	10.0	545	0
302AY retard.....	L—3.70	5.0	410	47
302BA retard.....	L—1.425	1.3	702	140
302E retard.....	L—3.52	1.1	514	35
320F retard.....	L—26.11	1.7	702	130
320G retard.....	L—19.62	10.0	653	1
320M retard.....	L—1.428	1.0	307	53
320N retard.....	L—18.0	2.0	54	0
320P retard.....	L—0.874	1.0	10122	41
1500A inductor.....	L—43.53	1.0	1883	121
1500B inductor.....	L—40.44	1.0	738	12
1507A inductor.....	L—283.9	7.0	560	10
	R—260.8	7.0		
1507C inductor.....	L—78.2	4.0	685	71
	R—2025	2.0		
1507D inductor.....	L—6.5	2.0	881	230
	R—166.6	5.0		
1507E inductor.....	L—23.9	5.0	828	88
	R—1655	5.0		

NOTE: Production includes product manufactured which did not fall within the "A" Limits.

the distribution specification. In the case of the 302BA and 320F retards, the lower yield of these coils was found to be due to machine variations which have been corrected. In the case of the 1507D inductors, an additional limitation is caused by the lack of fine enough machine pitch control to permit proper manufacture of this coil. A new winding machine having adequate control is being provided to eliminate this condition.

It has become apparent, as production has progressed, that the manufacture of coils, such as are given in Table III, is going to remain a small lot business. Therefore, in all probability, the great majority of the production will be shipped by the three-cell method, with only a few cases of enough production to warrant the use of the Batch or Continuous methods where advantage can be taken of the sampling procedures.

3.3 APPLICATION TO CAPACITORS

3.31 500-Type Molded Mica Capacitors

The 500-type molded mica capacitors have been one of the most difficult products to control within the requirements of the distribution

specification. This condition is inherent in the design and methods of manufacture of this type of product, as the process consists of an integrated group of manually-controlled operations. The following brief summary of these operations will give a better appreciation of the difficulties encountered in the control of a product of this type:

3.311 *Mica cut to size and silver coated*

- A. Inspected for visual and mechanical dimensional defects.
- B. Laminations silver coated by silk screen process and fired.

3.312 *Assembly*

A. Laminations stacked by count and interleaved with lead foil. In this state, the stack consists of laminations having multiple silvered areas.

B. Dip-sealed in Bi-Wax to hold pileup firmly together for subsequent operations and excess wax pressed out of pileup.

3.313 *Sectioning*

A. Section the multiple stack of laminations by punch and die to provide individual capacitor pileups.

3.314 *Preliminary stacking to capacitance range*

A. Rough adjust the individual capacitor pileups by adding or removing single laminations to meet a broad capacitance range which is capable of adjustment. Terminals are then attached by crimping.

3.315 *Adjusting*

A. Adjust by scraping silver from outer layer of mica until capacitance is within ± 0.2 mmf of specified value.

3.316 *Molding — Stabilizing*

- A. Mold in mica-filled molding compound.
- B. Stabilize by temperature cycling.

In a process of this type a number of individual operations combine to determine the final product capabilities. Control of distribution is difficult since many of these operations tend to substantially modify the distribution obtained in previous steps of the process. Figs. 9 and 10 show

the conditions encountered on two types of these capacitors which have different capacitance tolerances. Studies indicate that the percentage of the product falling within the cell limits at the present time will vary from 40 to 80 per cent depending on the capacitance value and tolerance. Thus an expected merchandise loss of 20 to 60 per cent falling outside of allowable tolerance exists which, in addition to the units classified but not eligible to ship due to poor distribution of product among cells, creates a condition resulting in increased cost of the product. It is expected, based on current methods and design, that the yield of the capacitors shipped within the requirements of the distribution specification will be between 20 and 70 per cent of the product manufactured. This represents a situation in which the specification tolerances, the product design, and the manufacturing process capabilities are not compatible, but because of the design needs it is still necessary to meet the distribution requirements. The use of the three-cell method makes this possible by the acceptance of relatively large shrinkages until improved manufacturing methods and materials now under investigation are introduced.

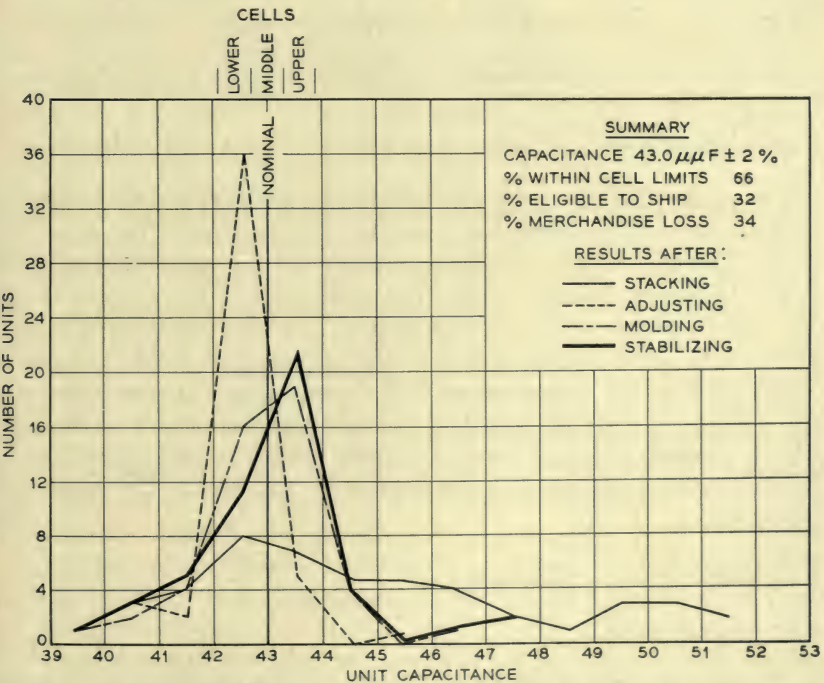


Fig. 9 — A summarized frequency distribution of a 43-mmF mica condenser at various stages of manufacture.

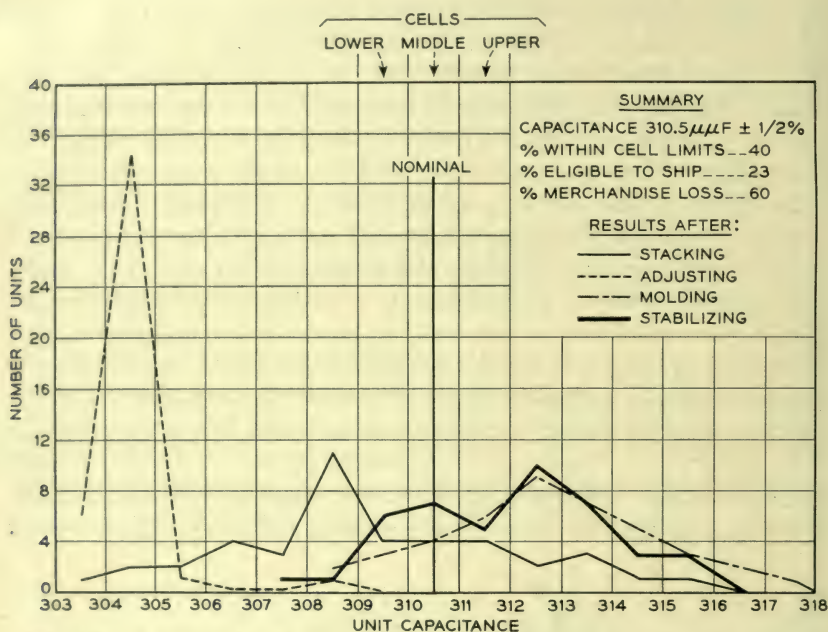


Fig. 10 — A summarized frequency distribution of a 310.5-mmF mica condenser at various stages of manufacture.

3.32 Quartz Disc Capacitors

The quartz disc capacitors are made by the addition of a silver coating to fused quartz discs which have been lapped to an exact thickness. The final adjustment for capacitance is obtained by the mechanical removal of small areas of the silver before the application of a protective finish. The nature of the manufacturing process and the quantities produced requires the three-cell method. A typical requirement is 9.8 ± 0.1 mmF which requires the measurement and classification of the product into cell widths of 0.067 mmF each. Capacitance measurements, including the use of reference standards, combined with careful adjusting procedures, have resulted in shipping approximately 99 per cent of the capacitors produced.

3.4 BORO-CARBON RESISTORS

3.41 Introduction

The stringent requirements for L3 carrier line amplifiers and impedance matching and balancing networks prevented the use of commer-

cially available composition type resistors due to their high temperature coefficient and lack of stability. Fortunately a new unit, known as a boro-carbon resistor was under development and could be utilized. The boro-carbon films employed in this unit have a maximum temperature coefficient of resistance of 0.01 per cent per °C compared to ordinary deposited carbon films which may have an average temperature coefficient of 0.03 per cent per °C.

Two types of boro-carbon resistors covering approximately 45 resistance values are being produced for the L3 system. These types are the 200A resistors, held to a tolerance of ± 3 per cent and the 200B resistor held to a tolerance of ± 1 per cent. Both types are available in resistance values from 5 ohms to 9999 ohms inclusive. Of the 45 values used in L3 equipment, approximately 40 are of the 200A type and 5 are of the more precise 200B type.

3.42 *Manufacturing Procedure*

The core of the 200-type resistor consists of a high grade alkaline earth porcelain of special composition. The resistance film is produced by placing the cores in a heated furnace containing a reducing atmosphere of hydrogen and injecting a combination of a hydro-carbon gas and boron trichloride into the furnace. Subsequently, cracking occurs depositing a thin film of boro-carbon over the surfaces of the cores. The resistance value of the film thus formed is dependent on the relative gas concentrations and the duration of exposure in the furnace. By suitable control, blanks having a resistance range from 5 to 250 ohms are produced. After removal from the furnace a band of silver paste, consisting of silver flake in a suitable binder, is applied to each end of the coated ceramic and baked. These bands form the contact surfaces for the terminals.

The resistors are next sorted into resistance groups preparatory to adjustment to the desired resistance value. Resistors having values between 35 ohms and 50 ohms are used to produce the higher resistance values between 250 ohms and 9999 ohms. The resistors falling outside the 35 to 50 ohm range are held for simple adjustment by rubbing. The method used to obtain higher resistance values is to cut a helix through the carbon film using a diamond cutting wheel. Resistance values may be raised by as much as 1000 times by proper selection of the pitch of the spiral groove.

After helixing, these resistors together with those initially outside the 35 ohm to 50 ohm range are fitted with a lead assembly at each end. Final adjustment of the resistance value is obtained by abrading the entire

coated surface uniformly to reduce the coating thickness. An adjustment of 15 per cent increase may be obtained by this means. After adjustment, resistors are given a protective coating, code marked and subjected to ten accelerated aging cycles over a temperature range of -50°C to 85°C .

3.43 *Use of Control Charts*

Since the manufacturing processes are essentially common within certain ranges of resistance it has been found possible to maintain control charts associated with the three-cell method for combinations of different values of resistance. In order to handle such combinations the characteristic under control is expressed as a percent deviation from the specified nominal. The control chart for averages then consists of averages of these percent deviations and the range chart is expressed in similar terms. This procedure permits a single chart to represent a large number of resistance values.

Control charts are used at two points in the manufacture of the 200-type resistors. The first chart is applied to the resistance value after adjustment either by rubbing, in the case of lower values, or by the combination of helixing and rubbing used for the higher resistance values.

Estimates based on preliminary studies predict that at half life this type of resistor will age upward by an average of 0.5 per cent. Manufacturing requirements for the nominal resistance values are set 0.5 per cent low in order to compensate for this condition.

The second chart is kept on the finished product after the accelerated aging cycles. This chart is also kept in terms of the percentage deviation from nominal. A comparison of the charts for the product after cycling, which represents the units which are candidates for stock, with those charts at the earlier resistance adjusting operation will show up changes in value due to processing. Appropriate changes in processing or adjustment bogies are made to obtain a more nearly centered value.

3.44 *Shipment of Product*

Although resistor production consists of a series of batch operations with each batch readily segregated and identified, the small quantities involved have resulted in the decision to package all of the product in accordance with the three-cell method.

The use of control charts after adjusting and after cycling has resulted in a more uniform product for the 200A resistors than would have been

possible by conventional manufacturing techniques. Initial production experience indicated that the "natural tolerance" of the 200A resistor is better than the ± 4 per cent "A" limit originally specified, and this limit was reduced to ± 3 per cent.

3.5 2504-TYPE TRANSFORMER

The 2504-type transformer, due to its critical function in the operation of the L3 amplifier circuit, has necessitated the use of completely new materials and methods not usually associated with transformer production. For instance the windings of this transformer are formed by diamond grinding the threads in fused quartz forms upon which a silver coating is bonded by a firing process. The grooves are then filled with electroplated copper, thus producing a winding intimately bonded to the quartz forms. Since the whole manufacturing process is unique and must be controlled in each small detail, application of statistical quality control procedures is essential.

Control of the leakage flux of this transformer, capacitance across the high impedance winding, and the capacitance to ground from the high potential terminal of its high impedance winding is of particular importance. To achieve the desired control of these characteristics, care must be taken to maintain a number of mechanical dimensions to tolerances far more precise than those on any transformer previously made by the Western Electric Company.

For example, the inner diameter of the quartz form, bearing the outer winding, is held to 0.7280 ± 0.0005 inches, while the thickness of the form between the inner diameter and the root diameter of the threads is held to 0.0310 ± 0.0005 inches. Dimensions of other component parts of the transformer must be held to comparable tolerances. To do this, it has been found necessary to keep all facilities — chemical, electrical and mechanical — under careful control at all times.

The facilities provided for the machining of the component parts are such that all critical dimensions which affect the final overall electrical characteristics of the transformer are held to approximately one-third of the design tolerances. By this means, it is expected the resulting transformers produced will meet the desired distribution. In the initial stages of production, records are being kept on 14 mechanical dimensions on each unit, checking them at each critical state of manufacture. The use of extensive mechanical tolerance controls on the components of this transformer is necessitated by the non-adjustable nature of its design. If controls were not applied, it would be impossible to obtain the close

electrical uniformity required by the end product. By means of special circuits and testing techniques developed for electrically measuring the characteristics of this transformer it has been found possible to indicate microscopic variations in secondary mechanical dimensions not directly under control.

The problem of bringing this transformer under control is not completely solved. Progress to date in the analysis of the data obtained indicates that the mechanical variations of the parts have complex inter-relations to the electrical requirements; however, manufacturing variations in the parasitic impedances mentioned above are being controlled to an order of magnitude better than on any transformer previously produced.

3.6 VACUUM TUBES

3.61 *General*

Three new vacuum tubes, the 435A, 436A and 437A have been developed for use in the L3 amplifiers. These tubes were described in detail in an article appearing in The Bell System Technical Journal of October, 1951.⁵ Two of them, the 435A and 436A are high transconductance tetrodes and the third tube, the 437A is a high transconductance triode. By applying the latest advances in design and in manufacturing techniques to tubes of conventional basic type, substantially higher levels of broadband amplifier performance have been realized.

The key to improvement in broadband amplification lies in an increase in figure of merit or transconductance to capacitance ratio. Figure of merit is a direct measure of the bandwidth over which the required level of amplification can be obtained. In general a given increase in figure of merit can be directly reflected in a wider transmission band which will provide more communication channels.

The higher transconductances and higher figures of merit obtained with the 435A, 436A and 437A over earlier broadband amplifying tubes are a direct result of advances in the art of manufacturing fine pitch grids of sufficient accuracy and rigidity to permit extremely close grid to cathode spacing. The basic objective is to provide a grid which can be placed very close to the cathode to act as a uniform potential plane controlling the cathode current without offering any physical obstruction to the passage of the current. This objective is approached by winding the grid with many turns of very small diameter wire.

The conventional method of grid manufacture consists of winding the grid lateral wire in a spiral around two side rods, usually of nickel. A groove is cut in the side rods at each point where the lateral crosses it

and the lateral is placed in this groove. The groove is closed by swaging to fix the lateral in place. In the L3 tubes, the grid lateral wire is approximately 0.0003" in diameter and the grid to cathode spacing is approximately 0.002". A ten per cent change in grid cathode spacing would result in a change in transconductance of fourteen per cent. It is important, therefore, that the control grid be maintained very accurately to insure proper grid to cathode spacing. A conventional grid made with such small diameter lateral wire would not be self-supporting in the length of span required and could not be made with the accurate control of diameter needed. Since the size of lateral wire used and the close spacing of adjacent turns preclude notching and swaging, some other method of holding the laterals in place must be employed. Accordingly, a new type of grid construction is used. This consists of making a supporting frame from two large side rods joined together by cross straps located at the ends of the grid proper. On this rigid frame the fine lateral wires are wound. This frame type of grid was described in previous articles.^{5, 6} In the L3 grids, the lateral wires have been bonded to the support rods by a glass suspension sprayed along the edge of the support rods. This glass glaze is sintered at a temperature of approximately 700°C to hold the laterals firmly in place. The earlier design used a gold brazing operation to secure the laterals. This brazing was done at a temperature of approximately 1070° C. The newer method at the lower temperature produces less stretching of the laterals and as a result higher residual tension is obtained. This is a distinct advantage in reducing noise and the possibility of grid to cathode shorts.

With the new method of holding the laterals in place, the grids are gold plated after the glazing operation is completed. Gold plating is necessary in order to minimize thermionic emission from the grid wires due to the closeness of the grid to the hot cathode and the possibility that active cathode coating material may be deposited on the grid during processing and operation.

3.62 *Distribution Requirements*

Distribution requirements have been placed on transconductance as well as the most critical inter-electrode capacitances, the input capacitance, C_{g1-kq2} , for the 435A and 436A tubes and the grid to plate capacitance, C_{g-p} , for the 437A. In the case of the 435A and 437A tubes, control of modulation is also required. Since transconductance is of primary importance, this characteristic has been selected as the one which governs shipment of product. Inter-electrode capacitance and modula-

TABLE IV — CHARACTERISTICS OF VACUUM TUBES STUDIED

435A and 436A	437A
Cathode current (I_k)	Plate current (I_b)
Grid-plate capacitance (C_{g1-p})	Plate resistance (R_p)
Plate-cathode, screen grid cap. (C_{p-kg2})	Heater-cathode cap. (C_{h-k})
Grid-heater cap. (C_{g1-h})	Heater-grid cap. (C_{h-g})
Heater-cathode, screen grid cap. (C_{h-kg2})	Heater-plate cap. (C_{h-p})
Heater-plate cap. (C_{h-p})	Grid-cathode cap. (C_{g-k})
	Plate-cathode cap. (C_{p-k})

tion are subject to the requirements of the Records method previously described.

In addition to the characteristics mentioned above, control charts were maintained on a number of other electrical characteristics at final test in order to determine process capabilities and to aid in setting final test specification limits. Among the characteristics that were studied were those given in Table IV. It was recognized at the beginning of production of these vacuum tubes that control of the test characteristics could not be achieved without similar control of the critical piece parts and processes going into the assembly of the tubes. Accordingly, control charts were started on the parts and processes, given in Table V, and have been maintained throughout the manufacture of the product.

Production of vacuum tubes represents a complex interlinking of many separate processes, each of which is essentially a batch type of manufacture. These batches vary considerably in size depending on the process or part involved. No part or process can usually be singled out as the controlling effect which would isolate one group of tubes from another. For example, a given lot of 100 tubes would probably be made from cathode blanks taken from a supplier's production run of 5 to 10 thousand parts. The cathode coating would be applied in batches of 50 to 200 parts. The control grid side rods used in making the grid frames would have come from a production run of perhaps 5000 parts and the

TABLE V — PARTS AND PROCESSES ON WHICH CONTROLS CHARTS WERE USED

Cathode blank	Outside diameter
Coated cathode	Outside diameter
Control grid support rod	Outside diameter
Control grid wire diameter	Unplated
Gold thickness	Control grid wire
Control grid	Minimum lateral resonance frequency
Screen grid	Minor axis, outside diameter
Plate	Inside diameter

fine grid wire used in winding the grid might be from a lot of 300 to 1000 meters etched to size at one time. The mounts are generally sealed into their glass envelopes in lots of 50 to 200 and pumped in lots of approximately 50 tubes.

Since the individual batches of parts and processing lots are of such different sizes, the final product does not result in groups of tubes that can be readily segregated into distinctive lots. As the tubes are assembled on a regular running basis the production is treated as being of a continuous nature and the Continuous Production method is applied. Mounts made from each week's production are identified by a serial number and, at test, samples from each week's production are used in determining conformance to the requirements of the distribution specification.

In the application of control charts to the L3 carrier vacuum tubes, the charts kept on the characteristics required by the test specification have been plotted against the *C*, *D* and *E* limits derived from the specifications. The charts kept on the parts and on other test characteristics have control limits derived from the process capabilities. In some cases these natural limits have been compatible with the original drawing limits and in others they have been at variance. Wherever incompatibility was established, an attempt was made to improve the process or where correlation studies justified the action, agreement was reached with the design engineers to increase the drawing tolerances.

3.63 *Application of Control Charts*

3.631 *Cathodes*

The cathodes used in the 435A, 436A and 437A are purchased from an outside supplier. A maximum variation from nominal of $\pm 0.0002''$ was specified for the minor axis outside diameter. These limits are tighter than for any previous similar cathode used, the narrowest limits heretofore being $\pm 0.0005''$. A control chart established on the first lot of cathodes produced for the 436A and 437A tubes indicated that a standard deviation (σ) of 0.00021" was obtainable. The 3-sigma limits thus derived of $\pm 0.00063''$ in addition to the displacement in the average ($\bar{X}$) for the lot of $+0.00026''$ represented a completely unsatisfactory condition. By selection of cathode blanks the nominal was reduced to 0.03121", approximately equivalent to the maximum permitted by the drawing. A much narrower range resulted from this selection. The standard deviation for selected cathodes amounted to 0.000066" corresponding to 3-sigma limits of $\pm 0.00020''$. However, this distorted distribution

was still unsatisfactory and attempts were made to obtain cathodes more nearly conforming to the design requirements. In the meantime a sizing operation was introduced in the shop to adjust the existing cathodes to size. By the additional sizing operation, the nominal value was reduced to 0.03102" almost exactly the desired value. However, the standard deviation of 0.00014" and 3-sigma spread of ± 0.00042 " were still twice the design intent.

A second run of cathodes was obtained in which the average measured 0.03099" and the 3-sigma limits were ± 0.00021 ". This substantially fulfilled the design requirements but made it imperative that the average be held precisely at nominal. To avoid differences between measuring instruments an agreement was reached to use a common type of electronic micrometer at both the suppliers' plant and the Western Electric Company. Control charts are kept by the supplier using a modification of the Continuous Production method outlined in the distribution specification. The application of this procedure has resulted in maintaining close control of the averages. Correlation studies on tubes manufactured with these cathodes, have shown that the limits for individual cathodes could be expanded to ± 0.00033 ".

3.632 *Coated Cathodes*

Three characteristics of oxide coated cathodes are particularly important in producing adequate emissive qualities, satisfactory life and the close spacing required in the L3 carrier tubes. These characteristics are weight of coating, density and coated dimensions particularly length and minor axis. In order to eliminate cathode blank variations from the measurements, dummy cathodes of known weight and precise diameter are included with each spray rack load of 41 cathodes. Measurements of gain in weight and of the increase in diameter due to coating are made on these dummy cathodes. The density can be calculated from these measurements. In addition to the measurements made on the dummy cathodes, control charts are kept on coated length and coated diameter on sample cathodes from each coating lot.

3.633 *Control Grid*

Close control of the minor axis for the frame type of control grid used in all three tube types is obtained by precision manufacture of the molybdenum side rods used in making the grid frame. The design specifications require a tolerance of ± 0.0001 " on the diameter of these rods. Adjustment of this diameter is obtained by tumbling the parts, which rounds

the ends of the rods at the same time to permit easy insertion of the grid in the support micas. Initial production was measured with a barrel micrometer read to the nearest 0.0001". The resulting data indicated a standard deviation of 0.00013". The 3-sigma limits of $\pm 0.00039''$ were four times the desired spread. Some portion of this spread was undoubtedly due to the measuring instrument which could not be read with sufficient precision. Subsequent production was measured with a Brown and Sharpe dial indicating barrel micrometer which was calibrated with standard gage blocks. Production measured under these conditions indicated the standard deviation to be 0.00007" for process limits of $\pm 0.0002''$. Fortunately, correlation studies indicated this tolerance to be acceptable. Control charts are kept on the product after the tumbling operation. The desired average is maintained by sampling the side rods for diameter several times during the tumbling operation.

The lateral wire used in winding the control grid is tungsten, etched to a diameter of $0.00029'' \pm 0.00001''$. The diameter of the wire is measured with a Bausch and Lomb metallurgical microscope equipped with a Filar eyepiece. The magnification used is approximately 450 $\times$. During initial production, it was not certain that wire of this size could be obtained consistently to a given nominal value and a chart was prepared adjusting the specified winding turns per inch to accommodate wire sizes from 0.00027" to 0.00032".

Control charts kept on the wire diameter have shown that it is possible to make wire consistently to a diameter of $0.00029 \pm 0.00001''$. Correlation studies have been made comparing grids of various wire sizes and corresponding turns per inch against the related tube characteristics. From these studies it was determined that with wire controlled to $0.00029'' \pm 0.00001''$, a single winding pitch for each tube could be substituted for the chart of turns per inch originally used. Subsequently, studies were made which enabled a common grid to be used in the 436A and 437A tubes.

After winding, the grids are gold plated. The desired increase in lateral wire diameter as a result of plating is 20 micro inches with limits of ± 10 micro inches. This measurement is also made with the microscope equipped with a Filar eyepiece.

Considerable difficulty has been encountered with measurement of wire diameter. Variations between operators have been encountered as well as differences between successive readings made at the same point on the wire. This difficulty has been minimized by careful training of the operators and by taking several readings at each point and averaging them.

The difficulties with instrumentation encountered with measurements of gold plating thickness make it more difficult to establish the standard deviation of the process. By determining the standard deviation of the measuring method used, it appears that a value of 4 to 5 micro inches may be realized. This corresponds to 3-sigma limits of 12 to 15 micro inches. Fortunately, correlation studies have established that this control has resulted in tube characteristics which have more than met the test requirement.

3.634 *Screen Grids*

The screen grids used in the 435A and 436A tubes are of conventional design. The tolerance for minor axis diameter is $\pm 0.001"$. The grids are first wound in strips on a mandrel which is shaped to the approximate dimensions of the desired finished grid. The strips are next degreased and given a preliminary heat treatment at 700°C for 5 minutes and then stretched longitudinally to straighten the side rods. The strips are then cut into individual grids and the loose grid lateral turns are trimmed from the ends. A heat treatment at 925°C for 15 minutes follows and finally the grids are sized, which consists of stretching the major axis on an expanding set of sizing blades to obtain the desired shape and size.

It was felt that it would be difficult to control the minor axis around a desired nominal since so many operations took place between the winding of the grid and the final sizing operation. Corrections for deviation of the center of the distribution usually consisted of slight changes in the amount of stretch imparted at the sizing operation and major corrections were usually attempted by a change in winding location on the tapered winding mandrel.

Process capability studies revealed a parent distribution of approximately $\pm 0.001"$ which is within the specified tolerance but provides no allowance for shift in average from one lot to another. Successive lots of grids processed in the same manner showed a considerable shift in center of distribution, amounting to as much as $\pm 0.0007"$.

Studies were then made to determine the important variables which needed to be controlled in order to minimize the shift in average value. It was found that for a given spool of grid wire, considerable control of the process average could be obtained by careful attention to two points in the processing. The grid minor axis size varied directly as the tension of the lateral wire was increased at the winding operation. By variation in tension, as much as $0.0005"$ shift in process average could be obtained. Secondly, when a more precisely controlled heat treating oven was used

which employed a thermocouple in each heat treating boat, it was established that the final minor axis dimension varied inversely with the heat treating temperature. Temperatures between 850° C and 1000° C could be used satisfactorily with a corresponding shift in process average of 0.0006". In Fig. 11(a) is shown the variation of grid minor axis with temperature and in Fig. 11(b) the variation of size with tension is indicated.

A procedure was then set up which resulted in a more uniform grid production. Each time a new spool of wire is used, a group of 25 grids is wound using a standard value of tension. These grids are processed through final sizing using 925° C for heat treatment and the standard sizing blades. The distribution of these 25 grids is then plotted. The position of the average is noted and correction is made for displacement of the average by a change in winding tension, a change in heat treating

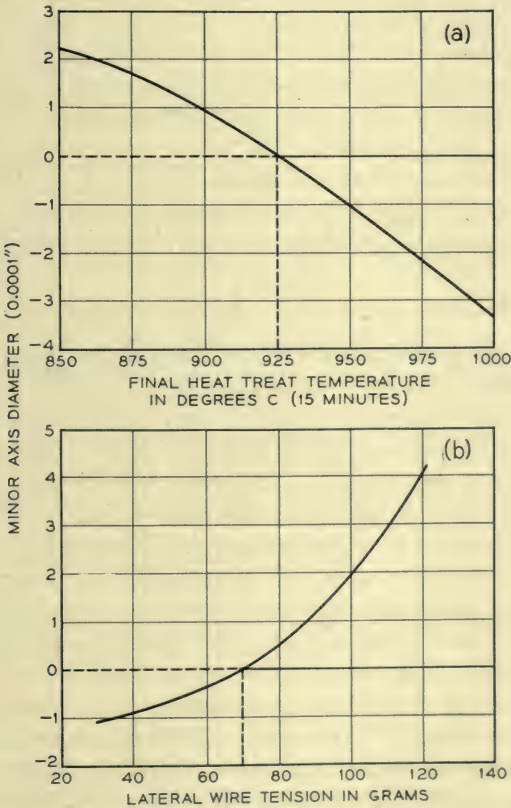


Fig. 11 — Control of grid minor axis in terms of temperature and tension.

temperature or a combination of both. Homogeneous lots of grids, are next run from this spool of wire and are processed up to the sizing operation using the same standard conditions. This test lot of 25 grids is run whenever a new spool of wire is introduced.

The first few grids from each lot are sized on the standard sizing blades and checked against a reset-run chart,⁷ a form of sequential analysis. The set-up is satisfactory if the distribution of the minor axis dimension is found to be within 1-sigma ($0.0003''$) of nominal. If grids pass the reset-run chart the balance of the lot is sized on the standard blades. If the sample does not conform to the reset-run chart, a final correction is made by substituting a larger or smaller set of sizing blades as required.

The distribution of the minor axis after sizing is plotted on control charts. A sample of 5 grids is taken from each tray of 36 grids and $\bar{X}$ and R plotted. If in control the lot is passed. If out of control, the lot is 100% inspected and packaged into three cells. The grids are then shipped to the mount assembly line in the usual 1-3-1 or 0-5-0 distribution. By application of the above controls of processing, variation in the process average from lot to lot has been reduced to $\pm 0.0003''$.

In order to keep an accurate record of the grid lots, relating the processes used to the resulting grid distribution, a special lot ticket has been developed. This ticket contains information such as specified tension and temperature along with the reset-run chart and $\bar{X}R$ chart. A ticket of this type accompanies each lot of grids through processing.

3.635 437A Plate

Since the 435A and 436A are tetrodes, the dimensions of the plate in these tubes are not nearly so critical as for the 437A tube, a triode. In the 437A, the plate assembly is made up of two sections welded together on a mandrel. An air press sizing operation is employed to control the inside diameter of the assembly. The drawing requirement for the plate assembly minor axis inside dimension was set at $0.0750 \pm 0.0005''$. Initial production was measured using a tool maker's microscope sighted on the edge of the plate assembly. Misalignment of the part under the microscope and burrs on the edge of the material contributed to variation in the measured values. In addition, the measurement was made at the edge of the assembly rather than the center which was desired. Results of measurements made with the tool maker's microscope are shown on the first portion of Fig. 12. The plate assemblies appeared to be oversize, although this was at variance with electrical test results on the tubes made from these parts. The standard deviation of $0.00101''$ and 3-sigma limits of $0.00303''$ were entirely too wide to meet the design intent.

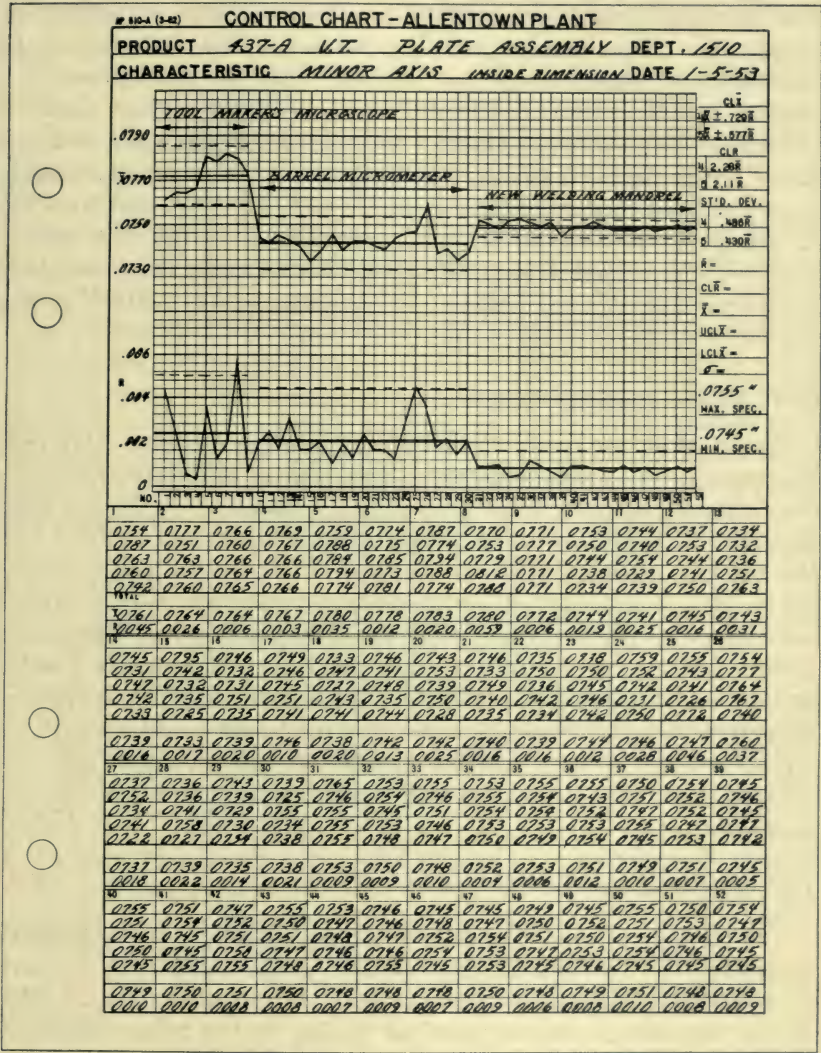


Fig. 12 — A control chart showing the results of using different methods of measuring the minor axis of plate assembly of the 437A vacuum tube.

A change was introduced in measuring technique at this point. The plate assemblies were measured at the center with a barrel micrometer. The thickness of the material was determined for each sample and subtracted to obtain the inside dimension. The point at which the new measuring method was introduced is indicated on Fig. 12. It will be noted that the average value shifted downward when the change in measuring

was introduced, although no change in the assembly process was made. The standard deviation of 0.00090" and 3-sigma limits of 0.00270" although better than obtained with the tool maker's microscope were still unsatisfactory. Based on the new average $\bar{X}$ for the process, a shift in welding mandrel was made directed at producing an average nearer to the 0.0750" value desired. The new welding mandrel was provided with heavy spring clips to hold the plate sections in position during assembly. A very uniform product resulted from this tool. The third portion of Fig. 12 represents production of plate assemblies using the new welding mandrel. The average $\bar{X}$ of 0.07495" was almost exactly that desired and the standard deviation of 0.00035" and 3-sigma limits of 0.00106" proved satisfactory and a change in the specification was introduced.

3.64 *Final Test Characteristics*

The controls kept on the critical parts as well as on many other parts of comparatively lesser importance have resulted in distribution of end requirements that have been well within the tolerances in the test specifications. In several cases the process average was found to differ from the nominal value specified. An analysis of the control charts enabled the design and manufacturing engineers to arrive at mutually acceptable adjustments of either the process or the test specifications. Where satisfactory performance would be obtained and the shift could be justified on the basis of control chart records, the specified nominal was adjusted to conform with the process average obtained. At the same time narrower limits, assuring a more uniform product to the L3 amplifiers, were adopted wherever it became evident that these were within the process capabilities.

4.0 CONCLUSIONS

The quality control methods described in this paper provide practical assurance that the selected characteristic of any component will have an acceptable distribution centered close to the specified nominal value. In this application, various statistical quality control techniques are an actual part of the product specification. The additional effort required in the application of such procedures is compensated for in part by the valuable assistance they provide in controlling the manufacture of the product. The procedures provide:

1. Means for determining compatibility of the product specification and the manufacturing process capabilities including the accuracy and stability of measuring facilities.

2. Useful techniques for improving manufacturing processes and for indicating the need of replacement or repair of worn tools and machines in the factory.

3. A device for promptly focusing attention on deviations due to changes in raw material or components.

All of these items are useful to both the design and manufacturing engineer in evaluating the relationship between the product design and the manufacturing and testing facilities.

The relatively small number of L3 amplifiers produced to date and the introduction of product design changes and modifications in manufacturing methods, which are inevitable during the early production period of a complex product of this nature, make it impossible to form final conclusions relative to the correlation between the distributions of related characteristics of the components and the final amplifier. Experience indicates, however, that these methods are of considerable assistance to the factory during the introduction of new designs of products of this nature and should, in addition, become a permanent and useful part of the production of products having such critical requirements.

5.0 REFERENCES

1. H. F. Dodge, B. J. Kinsburg and M. K. Kruger, The L3 Coaxial System — Quality Control Requirements, see pp. 943 to 968 of this issue.
2. W. A. Shewhart, Economic Control of Quality of Manufactured Product, D. Van Nostrand Co., Inc., New York, N. Y. (1931).
3. ASA War Standards Z1.1 — 1941, Guide for Quality Control; Z1.2 — 1941, Control Chart Method for Analyzing Data; and Z1.3 — 1942, Control Chart Method for Controlling Quality During Production. American Standards Association, New York, N. Y.
4. H. T. Wilhelm, Impedance Bridges for the Megacycle Range, Bell Sys. Tech. Jr. **31**, pp. 999 to 1012, Sept., 1952.
5. G. T. Ford and E. J. Walsh, The Development of Electron Tubes for a New Coaxial Transmission System, Bell Sys. Tech. J., **30**, 1103 to 1128, Oct., 1951.
6. E. J. Walsh, Fine-Wire Type Vacuum Tube Grid, Bell Lab. Record, **28**, pp. 165-167, April, 1950.
7. Dorian Shainin, Quality Control Methods — Their Use in Design — Part III, Machine Design, Sept., 1952.

Abstracts of Bell System Technical Papers* Not Published in this Journal

AHEARN, A. J.,¹ AND H. B. HANNAY¹

Formative of Negative Ions of Sulfur Hexafluoride, J. Chem. Phys., 21, pp. 119-124, Jan., 1953.

Negative ion production in SF_6 has been studied in a mass spectrometer; SF_6^- , SF_5^- , SF_4^- , F^- were identified. The SF_6^- and SF_5^- are produced in very large quantities by a resonance capture process at an electron energy of about 2 ev, and are formed in approximately equal amounts. At higher electron energies, the same capture occurs by a secondary process, in which low energy electrons released by other excitation and ionization processes suffer the resonance capture. Partial dissociation after electron capture accounts for the appearance of F^- and F_2^- ions below 16 ev. Above this energy, other processes will also produce these ions. Possible explanations of the primary resonance capture mechanism are discussed.

ANDERSON, F. B.¹

Gain and Phase Angle Measuring Set, Elec. Eng., 72, pp. 245, Mar., 1953.

Digest of paper "A 10-Cycle to 10-Megacycle Gain and Phase Angle Measuring set."

ARMSTRONG, C. A.²

Communications for Civil Defense, Elec. Eng., 72, pp. 218-222, Mar., 1953.

BECK, A. C., see S. E. MILLER.

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

¹ Bell Telephone Laboratories, Inc.

² American Telephone and Telegraph Company.

BECKER, J. A.,¹ AND C. D. HARTMAN¹

Field Emission Microscope and Flash Filament Techniques for the Study of Structure and Adsorption on Metal Surfaces, J. Phys. Chem., **57**, pp. 153-159, Feb., 1953 (Monograph 2073).

With field emission electron microscopes one can see the structure of the surface of a single crystal at the tip of a metal "point." The magnification is about 10^6 and the resolution about 20×10^{-8} cm. At 2800°K , the surface of *W* point is hemispherical. Only the 110, 111 and 100 regions consist of small flat planes. In fields of 50 million volts/cm. and 1200°K . these planes enlarge. The edges of the planes are seen to be in violent agitation. Hence surface atoms are mobile at temperatures above one-third of the melting point. Ba atoms show surface mobility at 400°K . on the 110 and at 800°K . on the 100 planes. With the flash filament technique one can measure the rate at which N_2 is adsorbed on a *W* ribbon at low pressures. After N_2 has been absorbed for minutes at a low temperature, the ribbon is flashed at 2300°K . The sudden rise of pressure is recorded with an ion gage and measures θ , the layers adsorbed. From a family of θ versus time curves one calculates s , the sticking probability. For $T = 300^\circ\text{K}$., $s = 0.55$ from $\theta = 0$ to 1.0. Then s decreases from 0.55 to about 4×10^{-4} for $\theta = 2.0$. These data yield an activation energy for the conversion of a molecule to two adatoms of about 100 cal./g. mole for $\theta = 0$ to 1.0 and 5000 cal./g. mole at $\theta = 2.0$. Other experiments yield 100,000 cal./g. mole for the heat of adsorption of 2 adatoms. The heat of adsorption of molecules is much smaller.

BOZORTH, R. M.,¹ AND R. W. HAMMING¹

Measurement of Magnetostriction in Single Crystals, Phys. Rev., **89**, pp. 865-869, Feb. 15, 1953 (Monograph 2074).

A simplified procedure is given for determining the five magnetostriction constants of a single crystal of a ferromagnetic cubic crystal. The crystal is cut as a disk parallel to a (110) plane, and strain gauges are cemented to the surfaces to measure strains in (001) and (111) directions. A magnetic field sufficient for saturation is oriented in 10° steps at various angles to the (001) direction, and magnetostriction is measured over a 90° range for each gauge. Each of the 18 data is then multiplied by suitable numbers, obtained by inversion of the strain matrix, to give the constants $h_1 \dots h_5$. The method is applied to a crystal of a 78 per cent nickel-iron alloy to determine the magnetostriction associated with spontaneous magnetization in the (111) direction: $\lambda_{111} = 2h_2/3 + 2h_5/9$, a quantity important in the "Permalloy problem." The constants are also determined for a single crystal of nickel.

BOZORTH, R. M.,¹ AND J. G. WALKER¹

Magnetic Crystal Anisotropy and Magnetostriction of Iron-Nickel Alloys, Phys. Rev., **89**, pp. 624-628, Feb. 1, 1953 (Monograph 2076).

¹ Bell Telephone Laboratories, Inc.

Single crystals of a number of iron-nickel alloys were prepared, and measurements made of the magnetic crystal anisotropy, and of the magnetostriction at saturation in different crystallographic directions, as dependent on the rate of cooling of the specimens after annealing. There is a large effect of the cooling rate on the anisotropy, for compositions near FeNi_3 , where atomic ordering occurs. There is a definite but smaller effect of cooling rate on the magnetostriction. The composition for highest initial and maximum permeabilities is nearly that for which λ_{111} , the magnetostriction in the direction of easy magnetization, is equal to zero.

DALTON, A. G.³

Practice of Quality Control, Sci. Am., **188**, pp. 29-33, Mar., 1953.

Statistical analysis of manufacturing processes has become a powerful tool of technology. An account of how its principles are now applied in the factory.

DRVOSTEP, J. J.,¹ AND A. W. LEBERT¹

Standardization of Rigid Coaxial Transmission Lines, Tele-Tech, **12**, pp. 78-79, Feb., 1953 (Monograph 2077).

DEHN, J. W., see R. W. Burns.

FELKER, J. H.¹

Arithmetic Processes for Digital Computers, Electronics, **26**, pp. 150-155, Mar., 1953 (Monograph 2078).

Special codes and arithmetical processes enable digital computers to perform rapidly many heretofore laborious mathematical tasks. Review of these processes serves as introduction to newcomers to field and review for veteran computer engineers.

GOUCHER, F. S.,¹ AND M. B. PRINCE¹

Interpretation of α -values in p - n Junction Transistors, Phys. Rev., **89**, pp. 651-653, Feb. 1, 1953 (Monograph 2068).

By the measurement of five parameters in several p - n junction transistors, viz., the conductivities and diffusion lengths of minority carriers in the emitter and base regions and the widths of the base regions, the current amplification factor α of the transistors has been computed from theory. Previous to this investigation two of the parameters associated with the thin p -layer had not been measured. The quantity α also was obtained independently by two alternate methods: (1) by measuring the collector-emitter current characteristic, and (2) by measuring the apparent quantum efficiency of the transistor as a two-electrode photocell with a floating base. The three determined values of α for each sample agree within the experimental error.

¹ Bell Telephone Laboratories, Inc.

³ Western Electric Company.

HAMMING, R. W., see R. M. Bozorth.

HANNAY, N. B., see A. J. Ahearn.

HARTMAN, C. D., see J. A. Becker.

HEWITT, W. P., see W. P. Mason.

HULM, J. K., see B. T. Matthias.

JONES, H. L.⁴

Approximating the Mode From Weighted Sample Values, Am. Stat. Assoc., J., **48**, pp. 113-127, Mar., 1953.

The weighted mean of ordered sample observations can be used to approximate the mode under favorable conditions, where the weights are determined from the first two terms in a Taylor expansion of the maximum likelihood estimate. Such weights are shown in Table 1 for the case where the sample is selected from a *t*-distribution with known kurtosis.

LEBERT, A. W., see J. J. Drvostep.

LEWIS, H. W.¹

Calibration of the Rolling Ball Viscometer, Anal. Chem., **25**, pp. 507-508, Mar., 1953.

MASON, W. P.,¹ W. H. HEWITT¹ AND R. F. WICK¹

Hall Effect Modulators and "Gyrators" Employing Magnetic Field Independent Orientations in Germanium, J. Appl. Phys., **24**, pp. 166-175, Feb., 1953 (Monograph 2079).

Three uses for the Hall effect in germanium crystals are described. These are (1) use of Hall effect probes in measuring magnetic flux, (2) use of Hall effect in crystals to produce a pure product modulator, and (3) use of Hall effect in germanium crystals to produce a nonreciprocal transmission. If the resistances are shunted around such gyrators, the transmission can be made zero in one direction and finite in the other. In all these applications, use is made of a crystal orientation for which the cross magneto-resistance effects are zero and the Hall effect constant does not vary with field by more than 2 per cent out to a flux density of 20,000 gauss. This orientation was located by making a phenomenological study of the magneto-resistance and Hall effect corrections for a cubic crystal and evaluating constants experimentally. Correction terms to fourth and fifth powers of the magnetic field have been obtained.

¹ Bell Telephone Laboratories, Inc.

⁴ Illinois Bell Telephone Company.

MATTHIAS, B. T.,¹ AND J. K. Hulm¹

Superconducting Properties of Cobalt Disilicide, Phys. Rev., **89**, pp. 439-441, Jan. 15, 1953.

A solid rod of cobalt disilicide was found to have a superconducting transition temperature close to 1.4°K, a critical field gradient of 146 gauss per degree at the transition point, an ice-point resistivity of 16.5 micro-ohm cm and a residual resistivity of about 16 per cent of the ice-point value.

MILLER, R. L.¹

Auditory Tests with Synthetic Vowels, J. Acoust. Soc. Am., **25**, pp. 114-121, Jan., 1953.

The results are given for a series of phonetic evaluation tests which were made by means of synthetically produced vowel sounds. By employing synthetically produced sounds, a number of the significant parameters could be varied in an independent and systematic manner without encountering the uncertainties and limitations of the human speech mechanism. The types of parameter changes which were investigated by this means were those of fundamental frequency or pitch, formant frequency and amplitude, and, finally, the number of formants important to a sound. The results of the tests indicate that all of these parameters are important in the evaluation of the sound. In particular, there is a shift in the phonetic evaluation which can be attributed to pitch alone.

MILLER, S. E.,¹ AND A. C. BECK¹

Low-Loss Waveguide Transmission, I.R.E., Proc., **41**, pp. 348-358 Mar., 1953 (Monograph 2080).

The circular electric mode in round metallic tubing becomes increasingly more attractive than the dominant mode from the standpoint of minimizing the waveguide size at frequencies above about 10,000 mc for the loss criterion of 0.25 db/100 feet. The circular electric (TE₀₁) mode also makes available a theoretical heat loss of 2 db/mile in waveguides less than 6 inches in diameter at frequencies higher than about 5500 mc. Increased transmission bandwidth, reduced delay distortion, and reduced waveguide size are factors favoring use of the highest practical frequency of operation. An increased number of freely propagating modes and smaller mechanical tolerances are the associated penalties. Experimental work has been carried out in the 9000-mc region using the TE₀₁ mode in a pipe about 5 inches in diameter. Transmission of 0.1-μ sec pulses had been observed over a distance of 40 miles. Mode conversion and surface roughness of the tubing walls result in observed losses which average about 50 per cent higher than the theoretical values for geometrically perfect, smooth-walled tubing. There is included a brief discussion of several problems unique to transmission in a multimode medium, including pure mode generation, mode filtering, the bend problem, and the effects of mode conversion on transmission loss and signal fidelity.

¹ Bell Telephone Laboratories, Inc.

NYE, J. F.¹

Some Geometrical Relations in Dislocated Crystals, *Acta Metallurgica*, **1**, pp. 153-162, Mar., 1953.

When a single crystal deforms by glide which is unevenly distributed over the glide surfaces the lattice becomes curved. The constant feature of distortion by glide on a single set of planes is that the orthogonal trajectories of the deformed glide planes (the *c*-axes in hexagonal metals) are straight lines. This leads to the conclusion that in polygonisation experiments on single hexagonal metal crystals the polygon walls are planes, while the glide planes are deformed into cylinders whose sections are the involutes of a single curve. The analysis explains West's observation that *c*-axes in bent crystals of corundum are straight lines. For double glide on two orthogonal sets of planes there is a complete analogy between the geometrical properties of the distorted glide planes and those of the "slip-lines" in the mathematical theory of plasticity. More general cases are discussed and formulae are derived connecting the density of dislocations with the lattice curvatures. For a three-dimensional network of dislocations the "state of dislocation" of a region is shown to be specified by a second-rank tensor, which has properties like those of a stress tensor except that it is not symmetrical.

O'CONNOR, S. F.³

Plating Room Waste Water Disposal, *Metal Finishing*, **51**, pp. 56-58, Feb., 1953.

OLSEN, K. M., see W. G. Pfann.

OWENS, C. D.¹

Analysis of Measurements on Magnetic Ferrites, *I.R.E., Proc.*, **41**, pp. 359-365, Mar., 1953 (Monograph 2075).

The unconventional behavior of permeability and core loss in the magnetic ferrites as compared to metals has led to a study of core-loss measurements. The relationships between the magnetic quality factor μQ and the characteristics of coils and transformers are developed, and the advantages of μQ as a parameter for the study and application of ferrites are discussed. A selected bibliography is given.

PFANN, W. G.,¹ AND K. M. OLSEN¹

Purification and Prevention of Segregation in Single Crystals of Germanium, *Letter to the Editor, Phys. Rev.*, **89**, pp. 322-323, Jan. 1, 1953.

¹ Bell Telephone Laboratories, Inc.

³ Western Electric Company.

PRINCE, M. B., see F. S. Goucher.

QUARLES, D. A.⁵

A.I.E.E. Progress, Elec. Eng., **72**, pp. 189-191, Mar., 1953.

In his address before the A.I.E.E. Winter General Meeting, President Quarles reviews some matters of current interest to the engineering profession, and to members of the Institute in particular.

RAE J. R.²

Microwaves from Coast-to-Coast, Gen. Elec. Rev., **56**, pp. 17-21, Mar., 1953.

STANSEL, F. R.¹

Transistor Equations, Electronics, **26**, pp. 156-158, Mar., 1953 (Monograph 2066).

Circuit gain and impedance characteristics are given in terms of transistor parameters for grounded base, grounded emitter and grounded collector configurations. Simplifying approximations are given where appropriate.

VARNEY, R. N.⁶

Drift Velocity of Ions in Oxygen, Nitrogen, and Carbon Monoxide, Phys. Rev., **89**, pp. 708-711, Feb. 15, 1953 (Monograph 2081).

The drift velocities of ions of the parent gas in oxygen, nitrogen, and carbon monoxide have been measured as a function of field strength to pressure ratio by techniques previously reported. Oxygen gave results similar to those in the rare gases reported previously. A log-log plot of drift velocity against E/p_0 in volts/(cm/mm Hg) starts with a slope near unity which gradually decreases to one-half at high values of E/p_0 . The mobility, extrapolated to zero field and atmospheric pressure is 2.25 cm²/volt-sec. Nitrogen and carbon monoxide both show a novel characteristic; the drift velocity first rises with E/p_0 but reaches a maximum and actually decreases, then finally resumes a more normal rise with E/p_0 as described for oxygen. It is believed that a high E/p_0 the drift velocity is characteristic of N₂⁺ ions and CO⁺ ions, respectively. At low fields the ion in nitrogen is believed to be N₄⁺. In CO the ion at low fields is believed to be CO⁺, with (CO)₂⁺ being formed at intermediate fields. The results are complicated by an additional ion which appears in the range of E/p_0 from 95 to 250 and which has a higher speed than the other ion. It is suspected of being C⁺.

¹ Bell Telephone Laboratories, Inc.

² American Telephone and Telegraph Company.

⁵ Sandia Corporation.

⁶ Washington University, formerly with Bell Telephone Laboratories.

WANNIER, G. H.¹

Connection Formulas Between the Solutions of Mathieu's Equation,
Appl. Math. Quart, **11**, pp. 33-59, Apr., 1953 (Monograph 2082).

The problem of connecting the various types of solutions of Mathieu's equation is solved by the introduction of a new parameter ϕ which is a function of the two equation parameters a and q . This quantity ϕ is introduced and enclosed between two very close analytic limits in section 2. In sections 3, 4, 5 precise definitions are given and information is collected for the three main types of functions which are to be connected. Section 6 contains the connection formulas. Section 7 reviews the status of knowledge achieved. Section 8 is an appendix on integral equations which are more general than those developed earlier in the text, but which appear to be of no use for the main purpose of this paper.

WALKER, J. G., see R. M. Bozorth.

WICK, R. F., see W. P. Mason.

¹ Bell Telephone Laboratories, Inc.

Contributors to this Issue

FRANK R. DICKINSON, B.S., Union College, 1927. Bell Telephone Laboratories, 1931-. Mr. Dickinson has been primarily concerned with the development of carrier telephony and is currently working on the L3 carrier system. During World War II he was engaged in the development of airborne radar bombsights. Member of Eta Kappa Nu.

H. F. DODGE, S.B., Massachusetts Institute of Technology, 1916; Instructor, Electrical Engineering, 1916-17; A.M., Columbia University, 1922; Engineering Department, Western Electric Company, 1917-25; Bell Telephone Laboratories, 1925-. Mr. Dodge was associated with the development of submarine detection apparatus, telephone transmitters and electro-acoustic devices until 1924. He then joined the newly organized Quality Assurance Department and as Quality Results Engineer has been concerned with the development and application of sampling inspection methods, quality control techniques, and quality rating plans based on statistical methods. His group is also responsible for the preparation of Laboratories Inspection Practices. Consultant to Secretary of War, 1942-44. Shewhart Medal, American Society for Quality Control, 1949. Award of Merit, American Society for Testing Materials, 1950. Member of American Society for Testing Materials; Fellow of American Society for Quality Control, American Statistical Association, and Institute of Mathematical Statistics. Chairman of the American Standards Association Committee Z1 on Quality Control, the American Society for Testing Materials Committee E-11 on Quality Control of Materials, and the Standards Committee of the American Society for Quality Control. Co-author with H. G. Romig of *Sampling Inspection Tables* (John Wiley and Sons, 1944). Member of editorial board of Industrial Quality Control.

R. D. EHRBAR, B.E., Johns Hopkins University, 1937. Bell Telephone Laboratories, 1937-. Mr. Ehrbar is in charge of equipment design and field operations related to the development of the L3 coaxial system. During World War II he worked on radar development for the Signal Corps. Member of the Institute of Radio Engineers and Tau Beta Pi.

C. H. ELMENDORF, III, B.S., California Institute of Technology, 1935; M.S., California Institute of Technology, 1936. Bell Telephone Laboratories, 1936-. Transmission Systems Development Engineer, 1952. Since joining the Laboratories, Mr. Elmendorf has been associated with the development of the coaxial repeater system and is currently in charge of the group responsible for the development of the L3 coaxial system. During World War II he participated in the development of microwave components and airborne radar systems. Member of I.R.E.

TUDOR R. FINCH, B.S., University of Colorado, 1938; M.S., University of Colorado, 1939. After joining the Laboratories, Mr. Finch spent two years in the study of relay contacts. From 1940-46 he developed networks and circuits for radar applications and more recently networks for the wide-band L3 coaxial transmission system. He is currently engaged in transistor network development for both military and telephone applications. Member of the Institute of Radio Engineers.

ROBERT F. GARRETT, Graduate Electrical Engineer, Johns Hopkins University, B.S.E.E. 1926. Western Electric Company, 1926-. Worked since graduation as an engineer and as an engineering supervisor on various assignments with the Western Electric Company in the Engineer of Manufacture organization. These assignments include the design of factory testing equipment, the supervision of various departments engaged in the engineering planning for field maintenance test sets, spiral four carrier and microwave equipment. Member of American Society for Quality Control.

R. SHIELS GRAHAM, B.S. in E.E., University of Pennsylvania, 1937. Mr. Graham has been principally concerned with the design of equalizers, electrical wave filters, and similar apparatus for use on long distance coaxial cable circuits for both telephone and television transmission. During World War II he designed circuits for electronic fire control computers for military use, and later developed methods for computing network and similar problems on a digital relay computer. Member of the A.I.E.E., Tau Beta Pi, and Pi Mu Epsilon.

E. I. GREEN, A.B. Westminster College (Fulton, Missouri) 1915, graduate student University of Chicago 1915-16, B.S. in E.E., summa cum laude, Harvard University 1921. Professor of Greek at Westminster College 1916-17; Captain Infantry Overseas Service 1917-19. American Telephone and Telegraph Company, Department of Development and Research, 1921-34; Bell Telephone Laboratories 1934-. From 1921 to 1940, and again from 1946 to 1947, Mr. Green was engaged in development work on toll transmission systems, principally in multiplex wire transmission. During the war, 1941 to 1945, he was responsible for

development of microwave test equipment for radar systems, radio monitoring and jamming equipment. In 1948 he was made Director of Transmission Apparatus Development, and in 1953 was appointed Director of Military Communication Systems. He is a Fellow of the A.I.E.E. and a Senior Member of the I.R.E.

ALEXANDER J. GROSSMAN, E.E., Rensselaer Polytechnic Institute, 1925. Bell Telephone Laboratories, 1925-. Transistor Network Engineer, 1952. Mr. Grossman has been engaged in the development of transmission networks since joining the Laboratories. Author of *Electric Wave Filters* in *Electrical Engineers' Handbook* (Pender and McIlwain, 4th ed.). Member of the Institute of Radio Engineers.

R. W. KETCHLEDGE, B.S., Massachusetts Institute of Technology, 1942; M.S., Massachusetts Institute of Technology, 1942; Bell Telephone Laboratories, 1942-. During World War II Mr. Ketchledge assisted in research related to infra-red detecting devices and in the development of sonar devices. After the war he spent two years working on the development of the Key West-Havana submarine cable system and from 1949-53 he was in charge of systems design for the L3 coaxial system. He was recently appointed Electronic Apparatus Development Engineer and is responsible for gas tube and storage tube development. Member of Sigma Xi.

BORIS J. KINSBURG, B.S., University of Southern California, 1926; M.A., University of Southern California, 1928. Southern California Edison Company, 1928-30; Bell Telephone Laboratories, 1930-. Since joining the Laboratories, Mr. Kinsburg has worked on research and development of broad band carrier systems using coaxial cable as the transmission medium. This includes amplifier development, study of cross-talk in coaxial conductors, requirement studies for coaxial equipment, equalization studies and television echo requirements and, currently, quality control studies of the L3 system components and reliability studies of the long-range submarine cable development. Member of the Institute of Radio Engineers, American Association for the Advance of Science, and Society for Social Responsibility in Science.

ROBERT H. KLIE, B.E.E., Polytechnic Institute of Brooklyn, 1945. New York Telephone Company, 1930-42; Bell Telephone Laboratories, 1942-. After spending two years in the Commercial Relations Department, Mr. Klie entered a group engaged in the development of radar systems. Since 1946 he has worked on coaxial systems development. Member of Tau Beta Pi and Eta Kappa Nu.

M. K. KRUGER, B.S., St. Lawrence University, 1920. Engineering Department, Western Electric Company, 1920-25; Bell Telephone Lab-

oratories, 1925-37; Western Electric Company, 1937-49; Bell Telephone Laboratories, 1949-. Mr. Kruger spent a few years as an instructor in the student assistant course and then became engaged in the design of filters, networks, and transmission testing equipment. He devoted twelve years at Kearny to the design of shop testing equipment for transmission apparatus. Since 1949 he has been concerned with the application of quality control methods for the L3 carrier system and, more recently, with general quality assurance work. Member of the American Society for Quality Control and Phi Beta Kappa.

G. H. LOVELL, B.S. in E.E., Texas A. & M., 1927; M.S. in E.E., Polytechnic Institute of Brooklyn, 1943. N. Y. Edison Company, 1927-28. Bell Telephone Laboratories, 1929-. From 1929 to 1948 Mr. Lovell was concerned with the development of crystal filters for carrier systems. Since then he has worked on the development of broadband amplifier networks.

LESTER H. MORRIS, B.S., College of the City of New York, 1935. Bell Telephone Laboratories, 1928-. Mr. Morris' first assignments were in the calibration of standard telephone instruments and, later, the development of acoustic impedance bridges. From 1930 to 1935 he conducted research in loudspeakers and since 1935 he has developed repeaters for coaxial systems. Member of Phi Beta Kappa.

JOHN W. RIEKE, M.S. in E.E., Purdue University, 1940. Bell Telephone Laboratories, 1940-. A member of the Transmission Systems Development Department, Mr. Rieke worked on television circuits before and after World War II. During the war he assisted in the development of radar indicators, and currently he is engaged in the development of broad band television transmission systems. Member of American Institute of Electrical Engineers, Tau Beta Pi, and Eta Kappa Nu.

T. L. TUFFNELL, Bell Telephone Laboratories, 1927-1941, Western Electric Company, 1941-. With the Laboratories, Mr. Tuffnell worked on terminal equipment for transatlantic telegraph and telephone service, and the design of vacuum tubes for carrier telephone systems. His work in the Western Electric Company has been chiefly concerned with engineering problems related to the manufacture of vacuum tubes.

R. A. WADDELL, B.S. in E.E., Rose Polytechnic Institute, 1936. M.S.E.E., Ohio State University, 1938; Westinghouse Electric and Manufacturing Company, 1939-1941; Western Electric Company, 1941-. Since January, 1941, has been in various phases of product and test planning on test sets, varistors and coils used in the telephone system. Engineer Resistances and Spool Wound Coils, 1951-. Member Eta Kappa Nu.

THE BELL SYSTEM
Technical Journal
DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXII

SEPTEMBER 1953

NUMBER 5

- Transmission Design of Intertoll Telephone
Trunks H. R. HUNTLEY 1019
- The Card Translator for Nationwide Dialing
L. N. HAMPTON AND J. B. NEWSOM 1037
- Development and Manufacture of Electroformed Conductor for
Telephone Drop Wire A. N. GRAY AND G. E. MURRAY 1099
- Semiconductor Diode Gates L. W. HUSSEY 1137
- A Review of New Magnetic Phenomena R. E. ALLEY, JR. 1155
- The Correlatograph — A Machine for Continuous Display of
Short Term Correlation W. R. BENNETT 1173
- A Statistical Study of Selective Fading of Super-High
Frequency Radio Signals R. L. KAYLOR 1187
- Acceleration Effects on Electron Tubes F. W. STUBNER 1203
- Arcing of Electrical Contacts in Telephone Switching Circuits
Part I — Theory of the Initiation of the Short Arc M. M. ATALLA 1231
- Polyethylene Insulated Telephone Cable A. S. WINDELER 1245
- Abstracts of Bell System Papers Not Published in this Journal 1257
- Addendum to "Delay Curves for Calls Served at Random" by
John Riordan 1266
- Contributors to this Issue 1267

SEP 30 1953

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

S. BRACKEN, *President, Western Electric Company*

F. R. KAPPEL, *Vice President, American Telephone
and Telegraph Company*

M. J. KELLY, *President, Bell Telephone Laboratories*

EDITORIAL COMMITTEE

E. I. GREEN, *Chairman*

A. J. BUSCH

F. R. LACK

W. H. DOHERTY

J. W. MCRAE

G. D. EDWARDS

W. H. NUNN

J. B. FISK

H. I. ROMNES

R. K. HONAMAN

H. V. SCHMIDT

EDITORIAL STAFF

J. D. TERBO, *Editor*

M. E. STRIEBY, *Managing Editor*

R. L. SHEPHERD, *Production Editor*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; Alexander L. Stott, Treasurer. Subscriptions are accepted at \$3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXII

SEPTEMBER 1953

NUMBER 5

Copyright, 1953, American Telephone and Telegraph Company

Transmission Design of Intertoll Telephone Trunks

By H. R. HUNTLEY

(Manuscript received June 12, 1953)

At the 1952 Minneapolis summer meeting of the A.I.E.E. a symposium on the nationwide toll switching plan went into such features as the fundamental plant layout, numbering plan, toll switching and automatic accounting equipments. The present paper is intended to round out this coverage of the plan with a further discussion of the transmission features.*

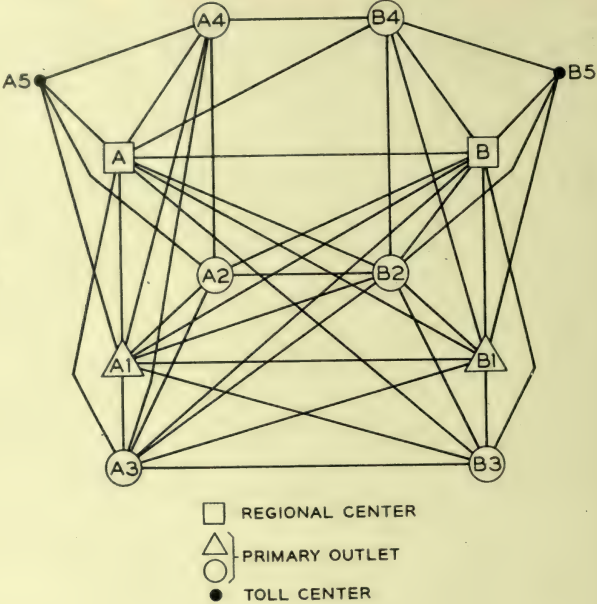
THE PROBLEM

In the new nationwide toll switching plan using switching machines the layouts of toll circuits and the routings of traffic will be quite different from that of the earlier plans which were based on manual switching. Individual calls can be switched so fast and cheaply that switching is no longer a limiting factor and circuits can be laid out and used in such a way as to obtain maximum economy with few, if any, limitations from the switching standpoint.

An example of these changes is given in Fig. 1 which shows in (a) the circuit groups which would be used to handle a given (assumed) flow of traffic on a manual basis and in (b) the groups which would be used to handle the same traffic on a dial basis. In (a) there are 44 different

* Trans. A.I.E.E., 71, Part I, Sept., 1952.

(a) RINGDOWN OPERATION



(b) NATIONWIDE TOLL DIALING

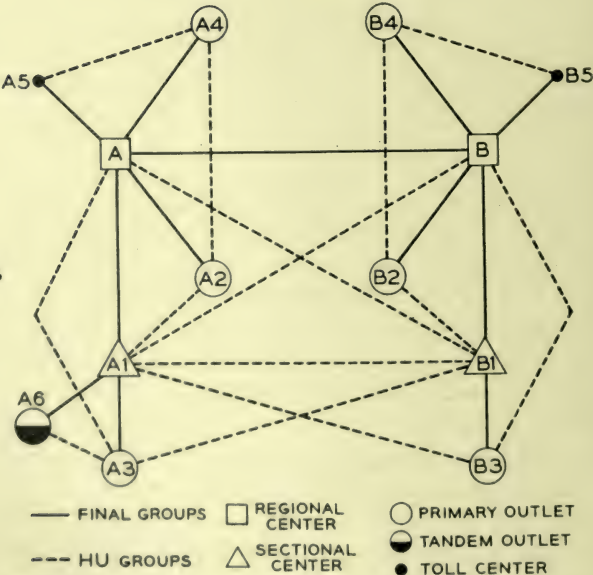


Fig. 1 — Typical intertoll trunk networks.

circuit groups and in (b) there are 26 circuit groups. More specific ideas regarding the effects of these differences can be obtained by considering how calls between specific centers (for example, A1 to B1) would be routed in the two plans.

From the transmission standpoint the principal impact of the new plan is that the situation will be changed from one in which as much of the traffic as practicable was handled over direct circuits with a minimum of switched traffic (circuits in tandem) to one in which two or more (up to a maximum of eight) circuits will be used in tandem on many calls and in which different numbers and make-ups of circuits may be encountered on successive calls between the same two telephones, as a result of the alternate routings employed with machine switching. This means that the losses of circuits must be low in order to provide adequate transmission on all calls and to avoid large differences in transmission on successive calls between the same two places.

The ideal method in such a situation would be to operate all circuits at zero loss since this would make the results independent of the number of circuits in tandem. However, the distances involved in the Bell System are so great that the propagation times, which affect echo, and the crosstalk between circuits require that even carrier circuits be operated at finite losses. Also, the plan must accommodate many voice frequency circuits on which the noise and singing conditions, as well as echo and crosstalk, may be more severe than on carrier circuits. The practical plan, therefore, is to:

1. Operate every circuit at the lowest loss practicable considering its length and the type of facilities used.
2. Assign circuits with different transmission capabilities in accordance with the parts they have to play in the operation of the over-all network.

The principal problem is to determine how low circuit losses can be made without getting into trouble due to one or more of the limitations mentioned above. This problem is complicated by the fact that the effects of these limitations are not directly proportional to circuit length or to the number of circuits in tandem. For example, if circuit (a) can be operated by itself at a loss of X db and circuit (b) can be operated by itself at a loss of Y db, the loss permissible when circuits (a) and (b) are switched together is less than $X + Y$. Ideally, therefore, each circuit should have a different loss in each different connection in which it is used. However, this is not practicable and a compromise must be adopted. This compromise provides that in some connections a particular circuit will operate at its lowest practical loss while in other connections higher losses will be employed to give over-all figures that will be ade-

quate from the standpoint of echo, crosstalk, etc. The general procedure is as follows:

1. When a toll circuit is switched to another toll circuit at both ends work it at a loss which is called "via net loss" (VNL).
2. When the circuit is switched to another at one end only (the other end being at the point of origin or destination of the call) work it at a loss higher than VNL by an amount which we shall call " S " (" S " being a generic term derived from the fact that it may be associated with switching pads — usually called " S " pads).
3. When the circuit is used by itself (i.e., the origin and destination of the calls are at the ends of the circuit) increase its loss by " S " again — that is, work it at VNL plus $2S$. This is known as "terminal net loss" (TNL).

Via net loss is, of course, to be the "lowest loss practicable" referred to above, and the next step is to establish methods of deriving VNL and of selecting the best value for " S ".

Since it would be a very complicated process to work simultaneously with all four of the limiting factors mentioned above, (echo, crosstalk, singing, and noise), the practical approach has been to select one of them as the basis of design and then check the results against the other three, modifying the final solution as necessary so that all four are kept under control. Since long experience indicates that echo is likely to be the most difficult and complex factor to control, it has been used as the starting point in the solution of the problem. As will be evident later, there are a large number of solutions possible from the echo standpoint and the one which has been selected has been affected to a considerable extent by the other factors.

The next part of the material in this paper is, therefore, devoted to an analysis of circuit design from the echo standpoint.

DETERMINING LOWEST PRACTICABLE CIRCUIT LOSS FROM ECHO STANDPOINT

The over-all objective is to have practically no cases in which objectionable echo will be observed by customers.

If circuits could be precisely adjusted to the requirements in each different connection the probability of echo would be the same on all connections and the computations would have been carried out on the basis of a very small probability — say, 1 in 10,000. However, losses can be changed only in discrete steps (S) so that in a very large proportion of cases the losses will be higher than are theoretically necessary.

Hence it seems sensible to compute the theoretical losses on the more liberal basis of 1 in 100, relying on the excess loss in most connections to reduce the over-all probability to the very small value desired.

The echo problem with which we are concerned is illustrated in Fig. 2. As shown there, part of the speech power which is being transmitted to the listener "leaks" across the hybrid (or four-wire terminating set) at the listener's end and returns to the talker. This is known as "talker echo." Actually, of course, some of the echo which returns to the talker can leak across the hybrid there and go back to the listener. This is known as "listener echo." However, with modern plant listener echo will not be important if talker echo is adequately controlled.

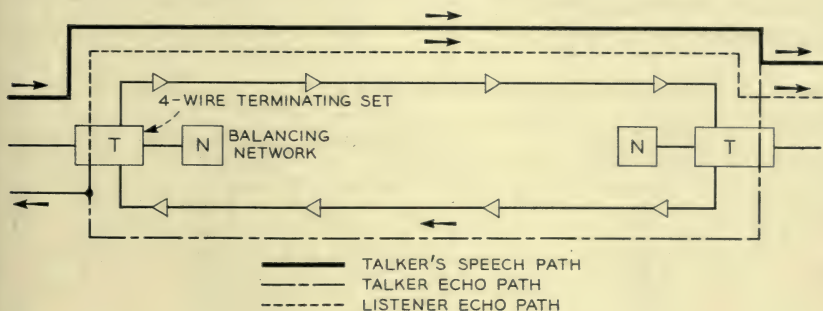


Fig. 2 — Echo paths.

Considering the effect of echo on the talker, if the elapsed time before the echo gets back to him is very short it is just like hearing his own voice through the sidetone in his own set, and unless it is very loud he doesn't notice it. If on the other hand, the elapsed time is long it sounds to him very much like the familiar acoustical echoes arising from physical obstacles. In extreme cases he may get the impression that the distant party is trying to interrupt him. The over-all effect of echo then depends on the following:

1. How loud it is — which in turn depends on how much loss there is in the echo path.
2. How long it is delayed before it gets back to him.
3. How easily he is annoyed by it (i.e., his "tolerance" to echo).

The factors involved are discussed in more detail in the following. For simplification four-wire circuits and four-wire switching are assumed; two-wire circuits and two-wire switching are treated as variations and are discussed later.

Tolerance to Echo

People vary greatly in their reaction to echo of given magnitude and delay and it is therefore necessary to treat them statistically, that is, what we need are the two statistical terms which are usually used to describe a mass of data, i.e., "average" and "standard deviation."

"Average" is simply the familiar algebraic average of the data — no talker is average but it is possible to obtain the average of a lot of talkers.

"Standard deviation" is a *number* which tells in general terms how the individuals spread out on both sides of the average. Usually: 30 to 35 per cent will be between the average and one standard deviation below the average; 30 to 35 per cent will be between the average and 1 standard deviation above it. At least 45 per cent will be between the average and two standard deviations below it and at least 45 per cent will be between

TABLE I

Round-Trip Delay (Milliseconds)	Loss in Echo Path Just Satisfactory to Average Observer
0	1.4 db
20	11.1 db
40	17.7 db
60	22.7 db
80	27.2 db
100	30.9 db

the average and two standard deviations above it. Very few, if any, will be outside of plus or minus three standard deviations.

Judgment tests under controlled conditions and with a number of observers (talkers), under conditions simulating connections to subscribers near the toll office, have given the basic data on these effects in Table I. (These data are slightly different from some published earlier because of recent reevaluations. Further studies are now under way and may indicate some further changes.)

An analysis of all the test data indicated that the observer judgments conformed fairly well with a normal law curve having a standard deviation (D_0) of 2.5. This means, for example, taking the 40 millisecond delay condition, that while on the average a 17.7 db loss was required in the echo path to make the echo just tolerable, some 30 to 35 per cent of observers could tolerate 2.5 db less loss. Another 30 to 35 per cent were sensitive enough to need 2.5 more loss for satisfactory echo condition. Practically no observer was so sensitive as to require $3 \times 2.5 = 7.5$ db more than 17.7 db, and practically none was so tolerant that he could permit 7.5 db less.

Terminal Return Loss

As shown in Fig. 2, with four-wire switching and four-wire type circuits the only source of echo is lack of perfect balance between the balancing network of the four-wire terminating set (hybrid coil or its equivalent) and the trunk, loop, and subscriber station connected at the customer side of the set at the distant terminal. The ratio of the amount of power reflected back into the hybrid coil to the amount which goes on toward the listener can be expressed as a loss in db. This "terminal return loss" in the echo range (approximately 500 to 2,500 cycles) has been found by tests and computations to have an average value of 11 db and a standard deviation, $D_t = 3$.

Round-Trip Circuit Loss

The round-trip circuit loss plus the terminal return loss is the total loss in the echo path. The round-trip circuit loss, i.e., the over-all loss which the intertoll trunk (or trunks) inserts in the echo path, is the sum of the losses in the east-to-west and west-to-east directions. If the circuit regulation were perfect, this loss would simply be twice the nominal one-way loss of the trunks — which is the thing we are looking for.

However, regulation is not perfect, and in order to determine what the nominal loss should be we must take into account the deviations from it which are certain to occur in practice. A considerable amount of experience indicates that these deviations can be treated statistically and considering some improvement in maintenance methods and procedures and a wider use of carrier systems with improved regulation, a standard deviation of $D_v = 2$ db for round-trip losses seems a not unreasonable assumption for the next few years.

Relationship Between Working Echo Net Loss and Round-Trip Delay

We now have all of the data we need to solve our problem — which as stated before is to find what VNL to assign to a circuit of given length and on a given type of facility.

Our first step mathematically is to combine the three statistical distributions we have been talking about — i.e., tolerance to echo, terminal return loss and the variations in round-trip circuit loss.

The combined standard deviation (D_c) of the three sets of distributions is the square root of the sum of the squares of the standard deviations of the individual distributions. The first two distributions are independent of the number of links, N , (assuming four-wire switching) but the distribution of circuit loss variations is a function of the number

of links. The mathematical expression for the combination of these three standard deviations is:

$$\begin{aligned} D_c^2 &= D_o^2 + D_t^2 + ND_v^2 \\ &= 2.5^2 + 3^2 + N2^2 \end{aligned}$$

From this equation Table II can be constructed.

In line with the principles stated at the outset, the mathematics will be worked out on the basis that 99 calls out of 100 will be free from echo; then margins will be added. The mathematics are as follows: (1) In order to meet 99 per cent of the cases, 2.33 standard deviations must be used, or: Avg. Rd. Trip Loss = Avg. Echo Tol. - Avg. Ret. Loss + 2.33 Std. Dev. (2) The average one-way loss is the loss to be assigned and is one-half the average round-trip loss.

TABLE II

No. of Links	Standard Deviation
1	4.4 db
2	4.8 db
3	5.2 db
4	5.6 db
5	5.9 db
6	6.3 db
7	6.6 db
8	6.9 db

Permissible average losses with several different numbers of links, based on this equation, are given in Table III. These data are for four-wire circuits and four-wire switching. It will be noted from Table III that for a given total round-trip delay the necessary increase in over-all loss for increasing numbers of links varies somewhat for different conditions but for the more severe cases it is about 0.4 db per link.

Because, as stated at the outset, the relationship between round-trip delay and permissible circuit loss is not linear, Table III can not be used directly in selecting the working net loss of a circuit to be used in switched connections. An example will make this clear:

(a) From Table III, the permissible loss of a circuit with 20 ms round-trip delay is 5.0 db.

(b) If this were used as the basis for designing the circuit, the loss of four such circuits in tandem would be 20.0 db whereas the table shows that four links with a total round-trip delay of 80 ms could be operated at 14.6 db.

The problem then is to find the best method of determining "VNL" and "S". As indicated earlier, this problem has a wide variety of solutions among which the best can be selected on a judgment basis. The process is as follows:

(a) In Fig. 3 a solid curve is shown giving the relation between working loss and round-trip delay for a single link, the information being taken from Table III.

(b) With the plant as it will be in the reasonably near future the round-trip delay on any connection without an echo suppressor will not exceed about 45 ms. This figure is based on a survey of geographical lengths, with some adjustment for the expected more extensive use of carrier and taking into account the "rules" (discussed later) for the use of echo suppressors.

TABLE III

Total Round-Trip Delay (Milliseconds)	Permissible Working Over-all One-way Loss (db)			
	1 Link	2 Links	4 Links	6 Links*
0	0.3	0.8	1.7	2.5
20	5.0	5.6	6.5	7.4
40	8.5	9.0	9.8	10.6
60	10.9	11.4	12.3	13.1
80	13.2	13.7	14.6	15.4
100	15.1	15.6	16.5	17.2

* With the switching arrangements which will be used, not more than 6 inter-toll trunks will be used in tandem without an echo suppressor.

(c) Then starting at any arbitrarily selected value of S , a straight line can be drawn from $2S$ (since there is S at each end) plus 0.4 (required to be added per link for variations) and intersecting the curve at 45 ms.

(d) In Fig. 3, three such straight lines are drawn, for $S = 1$, $S = 2$ and $S = 4$, which have slopes (in db per millisecond) about as follows:

S	Slope (db/ms)
1	0.15
2	0.10
4	0.016

(e) From these data, equations for VNL and TNL for four-wire circuits can be worked out in terms of round-trip delay, "d," thus:

S	VNL	TNL = VNL + $2S$
1	$0.15d + 0.4$	$0.15d + 2.4$
2	$0.10d + 0.4$	$0.10d + 4.4$
4	$0.016d + 0.4$	$0.016d + 8.4$

The slopes of the lines can be converted into factors (called "via net loss" factors — VNLF) in terms of db per mile by dividing twice the slope by the velocity of propagation in miles per millisecond. The product of VNLF and circuit length in miles plus 0.4 gives VNL. As an example, for K carrier circuits with a velocity of propagation of 105,000 miles per second, the via net loss factor for $S = 2$ would be $(2 \times 0.10) \div 105 = 0.0019$.

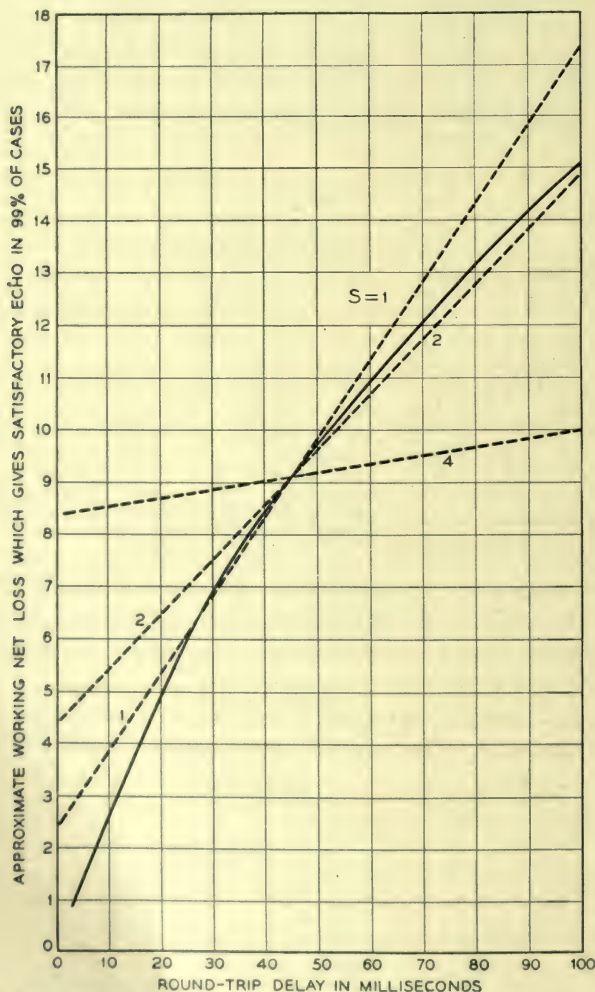


Fig. 3 — Approximate relationships between round-trip delay and permissible working one-way loss for an intertoll trunk from echo standpoint for four-wire circuits and four-wire switching.

For two-wire circuits the local echo paths at the repeaters make it impracticable to establish a straightforward relationship between over-all delay and echo performance. The via net loss factors for such circuits are approximations based on judgment and experience.

SELECTION OF VIA NET LOSS FACTORS AND S

From the foregoing, it is evident that the values of S and VNLF are interrelated and that there is a wide variety of possible relationships.

Fig. 3 shows that, up to fairly long delays, the lower the value of S , the lower the over-all losses at which the circuits can be worked from the echo standpoint.

Since the lower delay calls are much more numerous than longer delay calls, it is desirable to use as low an S as is practicable. However, in selecting a value, the other factors which have been neglected to this point — crosstalk, singing and noise — must now be taken into account and we must be sure that echo margin is now added. Each of these factors is discussed separately in the following.

Singing

The more extensive use of carrier reduces the importance of singing because voice frequency circuits are becoming shorter, thus eliminating the difficult singing problems associated with multi-repeater-section two-wire circuits. On the other hand, some of the conditions at circuit terminals may become more severe from the singing standpoint.

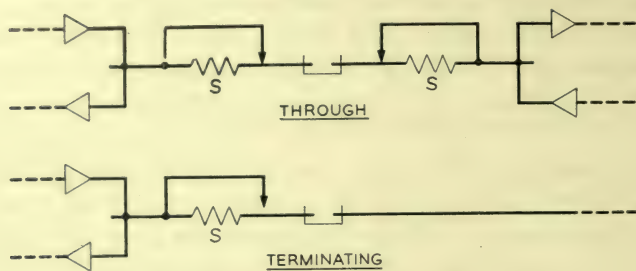
Studies indicate that over-all losses obtained with $S = 2$ are adequate to care for singing under most conditions but that if $S = 1$ were adopted, singing would be more important. With $S = 2$ the necessity for increasing circuit losses to avoid excessive danger of singing will probably be confined to a few open wire circuits having large discrete irregularities.

Noise

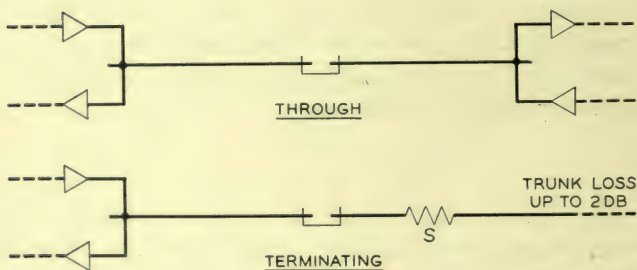
Noise is usually not a factor in the assignment of circuit losses. Carrier systems are designed so that under normal conditions the noise is low enough so that any desired loss can be used. If, in a specific case, noise in either carrier or voice frequency circuits is too high, the approach is to get rid of it by one or more of the means available.

Echo Margin

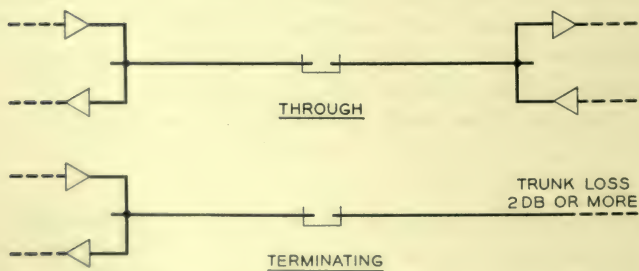
Reference to Fig. 3 will indicate that for $S = 2$ there is 2 db or more round-trip echo margin in all cases with round-trip delays less than the



(a) PAD IN INTERTOLL TRUNKS



(b) PAD IN TOLL CONNECTING TRUNK



(c) USING LOSS OF TOLL CONNECTING TRUNK

Fig. 4 — Methods of providing "S".

order of 30 or so milliseconds. With $S = 1$ there is little margin and with $S = 4$ there is excessive margin.

There will be very few connections with delays greater than about 30 milliseconds — which is roughly 1,500 miles of carrier — without echo suppressors. While the effect of the margin on probability of observing echo is difficult to compute quantitatively, it is estimated that with $S = 2$ this probability will be very small. Additional margin is also provided by the fact that on many connections the connecting trunk loss at the talker's end is greater than the loss used in the tests for establishing the echo tolerance curve.

Crosstalk

Analysis of many situations indicates, again, that with $S = 2$ the losses are about as low as are practicable in general from the crosstalk standpoint. With $S = 1$, they would be too low in many cases, and with $S = 4$, they would be unnecessarily high.

With $S = 2$, there may be a few cases where specific attention to crosstalk will be needed, particularly in open wire.

Final Selection

From the consideration of factors like the foregoing the value of 2 db has been selected for S on a judgment basis.

PROVISION OF S

The loss S can be provided in any of the following ways as appropriate. (See Fig. 4.)

1. As a switchable loss pad in the intertoll trunks.
2. As a fixed loss pad in the toll connecting trunk.
3. As part of the conductor loss of toll connecting trunks. This can be done only if the structural return loss of the connecting trunks against the balancing network is reasonably good.
4. If there is to be no switching to other intertoll trunks or to connecting trunks with more than 2 db loss, S may be provided simply by increasing the circuit loss by 2 db.

VIA NET LOSS FACTORS

Table IV lists typical VNLF's of Bell System intertoll trunk facilities for the condition $S = 2$. Typical losses at which circuits would be worked with $S = 2$ and with the via net loss factors tabulated in Table IV

TABLE IV — TYPICAL VIA NET LOSS FACTORS

Facility	VNLF (db per mile)	
	2-Wire Circuits	4-Wire Circuits
19H-88-50.....	0.03	0.014
19H-44-25.....	0.02	0.010
O. W. Voice.....	0.01	—
O. W. Carrier.....	—	0.0017
K or N Carrier.....	—	0.0019
L Carrier.....	—	0.0015
Radio.....	—	0.0014

are given in Table V. The advantages of high velocity, four-wire circuits (carrier and radio) are obvious from these tables.

ECHO SUPPRESSORS

Even if the intertoll trunk plant of the Bell System were all carrier the length of some connections would be so great that some method of controlling echo other than simply increasing circuit loss is desirable. Lower losses can be obtained on such connections through the use of an “echo suppressor,” an electronic device which under control of the talker’s speech currents places a high loss in the return path at the right time to intercept the return echo currents.

Echo suppressors perform very well so long as not many circuits equipped with them are connected in tandem and there is not too much time delay between them. With manual operation the switching is so limited that the chances of connecting circuits with echo suppressors in tandem are small and it has been practicable to apply echo suppressors on the basis of round-trip delay of the individual circuits. However, with dial operation it will be possible to establish connections which are long enough to require an echo suppressor but which are composed of circuits each too short to require an echo suppressor based on its round-trip delay. For example, an echo suppressor would not normally be used on a 500-mile carrier circuit, but if eight such circuits were connected in tan-

TABLE V

Type and Length of Trunk	VNL (db)	TNL (db)
50-mile N1 Carrier.....	0.5	4.5
50-mile 2-W H-88.....	1.9	5.9
200-mile 4-W H-44.....	2.4	6.4
500-mile K Carrier.....	1.4	5.4

dem giving a total length of 4,000 miles, an echo suppressor would be imperative, if over-all loss is not to be excessive.

It is not practicable to take care of this problem merely by reducing the delay time at which an echo suppressor is applied, since if this were done it is conceivable that eight circuits each with an echo suppressor might be connected in tandem. It has been necessary, therefore, to establish more or less arbitrary rules to insure at least one echo suppressor on long connections and to make it very improbable that more than two will be encountered. In general, these rules specify that echo suppressors will be placed on:

- a. All RC-NC circuits.
- b. All RC-RC circuits.
- c. On high-usage group circuits when the desired losses can not be met without them.

Our ideas as to when suppressors of Item c will be required may change with the trend from voice-frequency towards high-velocity carrier circuits. Experience will be a valuable guide, for it is not likely that an intolerable situation will build up overnight and without casting some shadow of coming echo; and the echo suppressor, being a discrete equipment unit, can be installed after it is found to be needed without appreciable lost motion or additional cost.

ALLOCATION OF FACILITIES

If the intertoll plant were homogeneous the over-all problem would be solved at this point — each circuit would be designed in accordance with the preceding and that would be that.

But the plant is not homogeneous — it consists of everything from loaded voice frequency circuits to circuits on microwave radio with VNLF's ranging from 0.03 to 0.0014. It is, therefore, necessary to allocate these facilities among different circuit groups in such a way that as far as practicable the higher performance facilities are used in the more demanding parts of the network.

As an aid to allocating facilities, charts like Fig. 5 are used. This chart shows ranges of losses within which circuits in different parts of the network are expected to fall. The losses shown there are exclusive of S which must be added, as indicated before, at both ends of each connection. It should be emphasized that these losses are not "limits" in the usual sense, neither are they attempts to divide up over-all losses among circuits. They simply help in allocating facilities in the non-homogeneous plant among different circuit groups. As the use of carrier

is extended, the plant will become more homogeneous and the need for such charts will gradually disappear.

Fortunately, from the transmission standpoint, while the machines will set up a wide variety of connections, the routing patterns will be rigidly controlled. Thus, it is practicable to know for each circuit group the maximum number of other circuits with which it can be used in tandem. The lower velocity circuits, two-wire circuits, narrow band circuits, etc. can (within practical limits) be allocated to groups which have relatively easy requirements.

TWO-WIRE SWITCHING — OTHER CONSIDERATIONS

Present views are that even the ultimate plan will involve two-wire switching at many points, mainly at the smaller switching points where

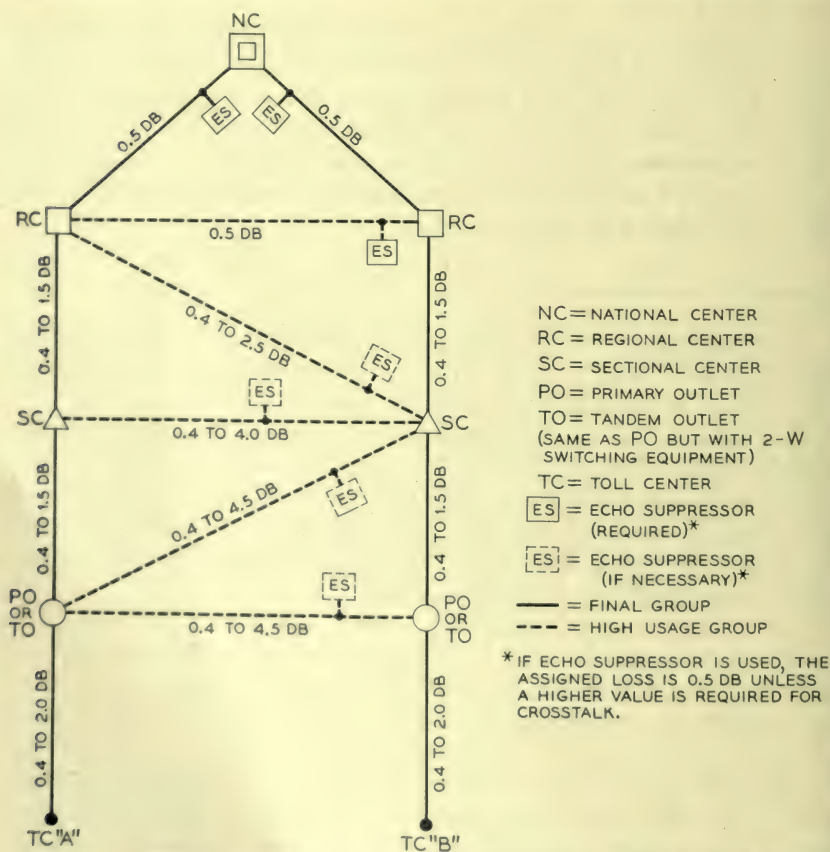


Fig. 5 — Intertoll routing pattern between two regions showing typical circuit groups.

the amount of traffic will not support the cost of complete automatic alternate routing features. This will cause some additional complication, for each such point introduces another source of echo due to the fact that the capacitance and resistance of the office cabling reduces the balance obtainable when two intertoll trunks are switched together. The effect of such switching on VNL's can be cared for by adding appropriate loss increments which will be small if a careful job of impedance matching is done and the distances from the toll terminal equipment to the switches is held within bounds.

No increment is added if the return loss for about 84 per cent of the circuits in the group is 24 db or more. These increments increase to about 0.2 db for a return loss of 20 db, 0.4 db for 18 db and so on. They are added to VNL of circuits between a two-wire switching point and a four-wire switching point and of circuits between two two-wire switching points. Impedance matching is usually accomplished by adding capacitance across the compromise network and in some cases across the shorter cable runs in an office.

All circuit losses referred to in this paper are 1000-cycle values, i.e., no allowance is made for the effects of noise and frequency distortion. Careful design, layout, and coordination of individual transmission systems are depended on to keep noise within proper bounds; and all new carrier systems going into the plant have transmitted bands wide enough to require no assignment of distortion transmission impairment (DTI). Circuits having excessive noise and those circuits with large DTI's are earmarked for improvement by any means that may come along. But beyond this, frequency distortion does not enter into VNL calculations since it can not be offset by reducing circuit losses without encountering trouble from the echo or other standpoint.

While we have considered only circuit design in this paper, it is evident that the success of the whole plan also depends on how closely circuit losses are maintained. This is important from two aspects.

1. The expected variations determine the allowance which must be made in the assigned loss. As indicated previously, it is expected that an allowance of 0.4 db per link will be adequate for the near future and it is hoped that as time goes on this figure can be reduced.

2. A more important factor is that unless circuit losses are maintained fairly precisely, large positive or negative excess losses can be accumulated on multi-switched connections. Avoidance of such large excesses is particularly important with dial operation since detection and avoidance of unsatisfactory transmission conditions by operators will be much less effective.

While the maintenance problem is at least as complex and difficult as the design problem, it is beyond the scope of this paper.

SUMMARY

To summarize the preceding discussion: For the particular conditions in the Bell System, a formula has been set up to give adequate approximations of the lowest practicable loss for practically all intertoll trunks as follows:

$VNL = VNLF \times L + A + B$, where:

VNL = "Via Net Loss" (db) of the trunk.

$VNLF$ = "Via Net Loss Factor;" i.e., a factor which depends on and is appropriate to the type of facilities used in the trunk.

L = Length in miles.

A = Design allowance for expected variations of circuit loss in service (0.4 db).

B = Amount to be added if two-wire switching is used; the magnitude depends on the passive return loss obtainable on such connections at the two-wire switching office.

At each end of the connection a loss of 2 db ($S = 2$) is added by appropriate means as discussed earlier.

CONCLUSION

Let it be emphasized that we have been talking largely of planning for the future in all that has preceded, for the switching plan as outlined is a growing thing and it will be a couple of years before much complex automatic alternate routing is done. And we would be very much surprised to escape growing pains and change of ideas as the plan develops. We are confident, however, that the plan is sound economically and transmission-wise; and flexible enough to adapt itself to further developments and experience.

ACKNOWLEDGMENT

As in most papers like this it would be prolix to mention all persons who took an active part in the preparation or in the development of the background data. But the author would be remiss if he did not call by name L. L. Bouton, who just prior to his recent retirement from Bell Telephone Laboratories, did much of the basic work on the mathematical concepts involved, on the simplification of these concepts for practical application, and on the re-evaluation of data that was required in these applications.

The Card Translator for Nationwide Dialing

By L. N. Hampton and J. B. Newsom

(Manuscript received August 24, 1953)

Nationwide operator and customer dialing requires the existence of a number of switching centers equipped with automatic systems having a much higher order of mechanical "Intelligence" than previous systems. One of the most important components of this new switching system is the Card Translator. Its function is to take the telephone address of a call and determine how to advance this call toward that address. This translator has to meet unique requirements in that it must accommodate a very large number of addresses; must provide a great amount of information for routing the calls; and must enable quick and convenient changes to be made in its stored information. It must also meet, of course, the normal basic requirements of reliability, economy, long life, etc. The fundamental principle of this translator is that of a card file, containing individual coded cards for each destination, with routing information recorded on each card. Whenever the routing information for a specific code is needed, the system selects the appropriate card, and reads the information by means of electronic circuits employing phototransistors and transistor amplifiers.

INTRODUCTION

The "Card Translator" was developed for the 4A toll crossbar system used at Control Switching Points (CSP) in the nationwide dialing network of offices. Although the many problems and conditions presented in the development and implementation of a nationwide dialing plan have been discussed in papers* by A. B. Clark, J. J. Pilliod, H. S. Osborne, W. H. Nunn, F. F. Shipley and others, it is necessary to restate some of these because of their effect on the translation problem and the features the card translator had to have in order to meet the nationwide dialing requirements. Translation is the process of converting the called destination code into information that is needed for the proper routing of the call.

* B.S.T.J., 31, pp. 823-882, Sept., 1952.

NUMBERING PLAN

For nationwide dialing it is necessary that each customer have a distinctive universal number. This numbering system is accomplished by dividing the country and Canada into about ninety numbering areas. Each of these areas is assigned a distinctive three digit code which, in order not to conflict with local office three digit codes, has either the digit "1" or "0" as the second digit. Within these numbering areas each local office will have a distinctive non-conflicting, name and number code. Since each customer in an office has a distinctive number, a corresponding distinctive nationwide universal number is thus provided. To reach a customer outside the local numbering area will require the dialing of three digits for the area code, three digits for the office code and four or five digits for the line number. Thus by dialing ten or eleven digits a connection can be made to any customer anywhere within the country and Canada. Fig. 1 shows the present numbering area code assignments for the United States and Canada.

TOLL LINE SWITCHING NETWORK

A second requirement for nationwide dialing is the provision of about 70 strategically placed automatic switching toll offices called control switch points (CSP) throughout the United States and Canada. The switching system used at each of these 70 or so CSP offices have several new features as follows:

1. Six digit translation.
2. Ease in changing and adding routings.
3. Automatic alternate routing.
4. Code conversion.
5. Storing and sending forward digits as needed.

BASIC SWITCHING ARRANGEMENT

In the CSP offices the transmission paths are established through crossbar switches mounted on the incoming and outgoing link frames as shown in Fig. 2. The setting up of the connection through these switches and the linkages is controlled by equipment common to the office which is held in use only long enough to set up each connection.

The major items of common control equipment are the senders, decoders, markers and card translators.

The sender's function is to receive and register the digits of the called destination, to transmit the area and, if required, the office codes

to the decoder, and then, as directed subsequently by the marker, to send digits ahead as may be required.

The decoder's function is to receive the code digits, either 3, 4, 5 or 6, from the sender and to submit them to the translator for translation and to make selections of alternate routes as required to route the call to the destination. The decoder also gives instructions to the sender and marker to enable them to carry out their functions.

The marker gets access to an outgoing trunk group through the trunk block connector and selects an idle outgoing trunk in this group, then chooses an idle linkage between the incoming and outgoing trunks, operates the crossbar switches to close the transmission path, and gives the sender information for pulsing ahead as may be required.

The operation of these common control circuits is briefly as follows. On the arrival of a call the incoming trunk is connected to a sender through the sender link frame. When the code has been registered the sender makes connection to the decoder through the decoder connector circuit. The decoder passes the code to the translator for translation. There is a translator associated with each decoder which contains the three-digit cards for the local offices and area codes, also perhaps some six-digit code cards. However, most of the six-digit cards (there may be several thousands of them) will be in a group of foreign area translators used in common by all decoders. The decoders obtain information from the area code card for selecting the particular one of these translators. Connection to them is made through an appropriate translator connector. The translator gives information for the selection of an outgoing trunk and passes other information for routing the call both to the decoder as well as to the marker. The marker proceeds with trunk test and the operation of the switches to establish the connection. When the proper information has been given to the marker, then the decoder and translator release from that call. When the marker has given information to the sender for routing the call and has established the talking connection, the marker releases. The sender releases as soon as it has finished pulsing ahead. In this way these common control units handle many calls in rapid succession.

SIX DIGIT TRANSLATION

This feature is needed in nationwide dialing because from a particular CSP to points in another numbering area there may be several routes or trunk groups. To reach a point in such an area it is necessary that the office code as well as the area code be translated to select the route to



* OKLAHOMA AND ONTARIO ARE TO BE DIVIDED INTO TWO AND THREE NUMBERING PLAN AREAS RESPECTIVELY. THE NEW AREA CODE NUMBERS HAVE NOT BEEN ASSIGNED

Fig. 1 — Nationwide toll dialing

the desired point in the area. To other areas there may be only one route and in this case the area code will suffice.

The provision of facilities for the translation of six digits greatly affected the design of the switching system for the nationwide dialing CSP offices. It led to the development of the basically new card translator. In previous toll common control systems translation is done by means of relays. The code digits, never more than three of them, resulted in the operation of groups of relays in certain combinations and led to the eventual operation of a route relay for the particular combination of code digits. This route relay with cross-connecting facilities from its contacts is used for the identification and selection of trunk groups and other information as might be required for the particular code routings. To change a routing with this system of translation required the removal and reconnection of many cross-connections.

With the nationwide dialing plan in operation, routing changes or opening of new offices in one part of the country will necessitate translator changes in many offices, some of them far removed from the scene of the event that forced them to be made. The changes in any one CSP may, therefore, be frequent under certain conditions, and to make them by running cross-connections would be cumbersome and expensive. The new translator uses punched cards instead of relays making it possible to effect changes by the simple process of removing old cards and insert-

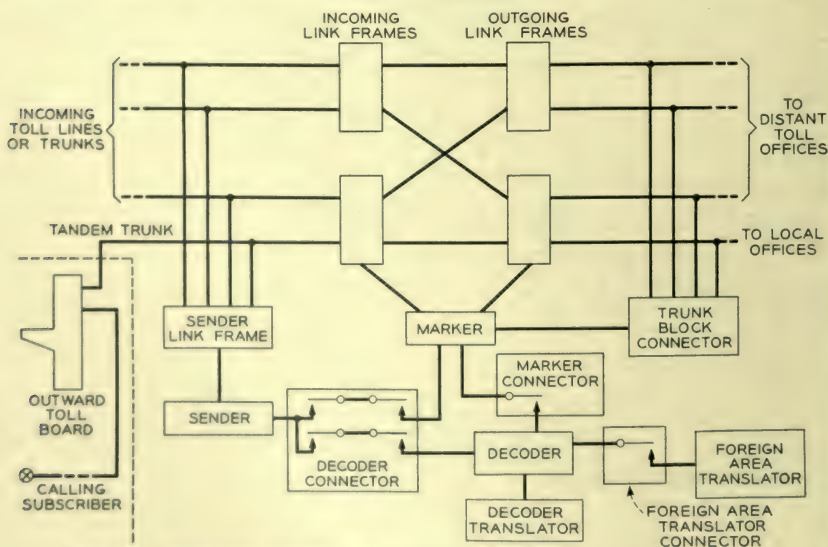


Fig. 2 — Schematic diagram of crossbar switching system for CSP's.

ing new ones in the machine. This can be done in a very short time and requires less out-of-service time for the equipment.

CARD FILE

A card is provided for each area code and also one for each office code that must be translated in a particular CSP, the cards representing destinations. The cards are lined up in a box as in a filing drawer with tabs along the bottom of the cards resting on select bars which run the length of the box. It is by operating the select bars in combinations, depending upon the code, that the particular card for the destination is selected. Each card, as shown in Fig. 3, has tabs, one for each select bar along the bottom edge. The information presented to the card translator for the selection of a card is in the form of code digits on a two-out-of-five basis. Each card is coded by removing all of the tabs except those that represent the particular combination of select bars for the particular code. When a code is presented to the translator, a combination of select bars corresponding to the code is lowered and the card having all tabs removed except those that were resting on the lowered selects bars will be selected while all other cards will remain in their normal positions.

The groups of tabs labeled, A, B, C, D, E and F (Fig. 4) are for the six code digits. For each digit, two tabs remain, since the digits are

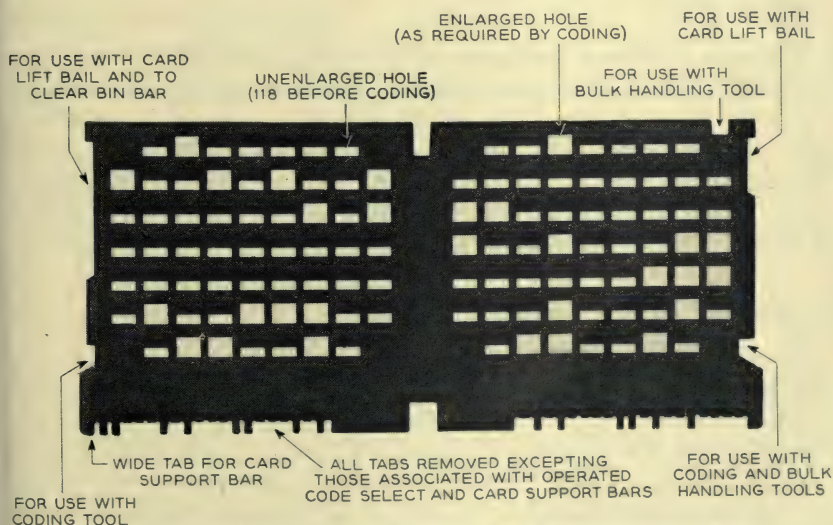


Fig. 3 — Typical coded card as seen from the phototransistor side.

registered in the sender on a two-out-of-five basis and the leads from the sender will cause the select bars to be operated. If the card represents an ordinary three digit code, all the tabs will be cut off except two each for the A, B, and C groups of tabs, and, for reasons that will be discussed later, two of the four CG tabs, card group, will also be removed. In addition, either the VO or NVO tab may be removed. The VO and NVO tabs are used when the group of toll lines over which the call will be routed is divided into one sub-group of a transmission grade suitable only for terminal traffic (NVO, meaning "not via only") and another subgroup for either terminal or switched-thru traffic (VO meaning "via only"). If the card represents a six-digit code, two tabs will be left in each of the six code digit positions and a different pair of card group tabs will be used.

The cards, then, are in different groups and are selected by combinations of code select bars together with the card group select bars. As noted there are four of these card group bars, CG0, CG1, CG2 and CG4. They are used in the six possible combinations, two at a time as follows: For the regular three-digit code cards CG1 and CG4, for the regular six-digit code cards CG2 and CG4; the other four combinations CG0 and CG1, CG0 and CG2, CG1 and CG2, and CG0 and CG4 are used for selecting the four groups of alternate route cards, which may be of the three-digit or six-digit variety.

CODE CAPACITY

The card translator by means of the code select bars and card tabs provides facilities for a great number of different codes and routings. There are 40 select bars provided, 36 of these are used in the combinations as has been described, two are reserved for possible future use and the remaining two are used for aligning the cards as will be discussed later. The total possible card code combinations is sufficient for growth in nation-wide dialing for the foreseeable future.

TRANSLATOR CARD CAPACITY

In some CSP offices there will be many thousands of cards for the destinations to be reached. It was not mechanically practicable to design a single translator capable of accommodating all these cards and moreover it would not have been economical, particularly for many of the smaller CSP offices. Also for service hazard reasons and to provide for the simultaneous translation of several calls, needed to handle traffic during heavy load periods, more than one translator is required in a

CSP. Considering these factors, it was found desirable to design the card translator to accommodate over a thousand cards.

TRANSLATORS

Since several translators are needed in a CSP, for further economy and consistent with service hazards, the translators are segregated according to the groups or kinds of cards they contain. These are as follows:

Decoder Translators

These translators contain the local three-digit office code, the three-digit area code, the alternate route code and, where space permits, some of the high usage six-digit foreign area code cards. Several of these translators are furnished, one for each decoder, depending upon the volume of traffic handled by the particular CSP.

Foreign Area Translators

These are furnished, one or two per office (maximum 19), for each 1000 or so foreign area code cards. These translators are in a common pool and the particular one is selected when needed for the six-digit code to be translated.

Decoder Foreign Translators

If there are sufficient high calling rate, six-digit, foreign area cards, to justify it, these translators may be provided, one per decoder.

Emergency Translator

Provisions are made so that the emergency translator can be substituted for any other translator by changing the cards of the translator in trouble to the emergency translator.

TRANSLATOR OUTPUT

The translator output information required for a CSP office for nationwide dialing must be very extensive to accommodate the many varieties of routes over which calls must be completed. The need for automatic alternate routing, code conversion and the storing and sending forward of digits also affects the need for increased translation output. The output is provided by means of the 118 holes in the face of each card.

PRETRANSLATION										OGT APPEARANCE				TRAFF. SEP. PC				TRK. GRP. PC & OF			
NCA CA4 CA5 CA6										IT TC ITC				TS0 TS1 TS2				TP0 TP1 TP2			
TRANSLATOR BOX NUMBER														CLASS							
INDI HB BTO BT1 BU0 BU1 BU2 BU3 BU4 BU7																					
AREA CODE CONTROL										ALTERNATE ROUTE PATTERN NUMBER											
NAC AC AHA AFA										ART0 ART1 ART2 ART4 ART7				CLT0 CLT1 CLU0 CLU1 CLU2 CLU4 CLU7				CDC7			
ROUTING INSTRUCTIONS										CONT & DIGIT CONTROL				ARU0 ARU1 ARU2 ARU4 ARU7				IND2			
RIO RI1 RI2 RI4 RI7																					
CODE CONVERSION																					
CCHN CCTN CCUN CCH0 CCH1 CCH2 CCH4 CCH7 CCT0														CCT1 CCT2 CCT4 CCT7 CCU0 CCU1 CCU2 CCU4 CCU7							
VAR. SPILL CONTROL										TRUNK BLOCK CONNECTOR				TRUNK BLOCK							
NSK SK3 SK6										TCU0 TCT0 TCT1 TCT2 TCU0				TCU1 TCU2 TCU4 TCU7 TB0 TB1 TB2 TB4 TB7							
GROUP START														GROUP END							
GST0 GST1 GSU0 GSU1 GSU2 GSU4 GSU7														GET0 GET1 GEU0 GEU1 GEU2 GEU4 GEU7							
A B C										D E F				CG							
0112147 0112147 0112147 0112147										0112147 0112147 0112147 0112147				0112147 0112147 0112147 0112147							

Fig. 4 — Card layout as seen from the exciter lamp side.

By enlarging these holes in combinations the output information for the particular route is obtained.

As seen in Fig. 4 the top holes beginning at the left of the card are used for "pretranslation" purposes. The senders are not provided with facilities which enable them to predetermine when to present for translation the first three digits received or when to wait for more than this number as when six-digit translation is required. Therefore the sender always requests translation when the first three digits are received. So for six-digit calls, the sender must be informed to disconnect from the translator after three digits have been received and wait for six digits. The CA6, (come again six) hole in the card is used for this purpose. The CA4 and CA5 holes are used similarly for calls to certain four- or five-digit operator codes, informing the sender to apply again for translation with four or five digits. The NCA (No Come Again) hole is used for three-digit calls.

The "OGT" holes are used to inform the common control equipment on which train of switches the outgoing trunk appears to enable the associated circuits to select the proper switching train. The remaining holes on the top line are for controlling operation of traffic meters.

On the second line, the TRANSLATOR BOX NUMBER holes are used on area code cards to indicate which translator contains the particular cards for the called area when six-digit translation is required. The INDI hole on the second line and the IND2 hole on the fourth line which commonly are referred to as index holes are never enlarged. They serve as an indication that a card has dropped and that all is ready for translation output detection. These index holes also aid in trouble detection in case of light failure, for routing of certain calls where cards are deliberately omitted and for calls where a blank code was dialed in error.

The CLASS holes are used for indicating the type of outpulsing and the kind of signalling channels used on trunk groups out of the office.

The AREA CODE CONTROL holes on the third line are used for determining the number of digits to be transmitted forward to the next office and for supplying undialed code digits needed primarily in connection with automatic alternate routing. The ALTERNATE ROUTE PATTERN NUMBER holes are used for the selection of the series of alternate routes to be used.

The holes on the fourth line are for making proper disposition of calls when all trunks are busy and to inform the associated circuits how many digits should be received for the particular code.

The CODE CONVERSION holes on the fifth line are used to supply the sender with information as to the outpulsing of certain arbitrary digits

as may be required through step-by-step toll trains. Facilities are provided for the outpulsing of one, two or three digits as may be required.

The VARIABLE SPILL CONTROL holes on the sixth line inform the sender when to pulse forward all digits as received, or to omit sending the first three or six code digits.

The remaining holes on the card define the location on the switching frames having the desired outgoing trunk appearances. The notches around the outer edges of the card are for proper positioning of the card in the stack and for card removal purposes as will be discussed later.

CARD OUTPUT DETECTION

The enlargement of the holes in the face of the card to obtain the translation output as previously stated is recognized by means of modulated light beams falling on phototransistor detectors. With all of the

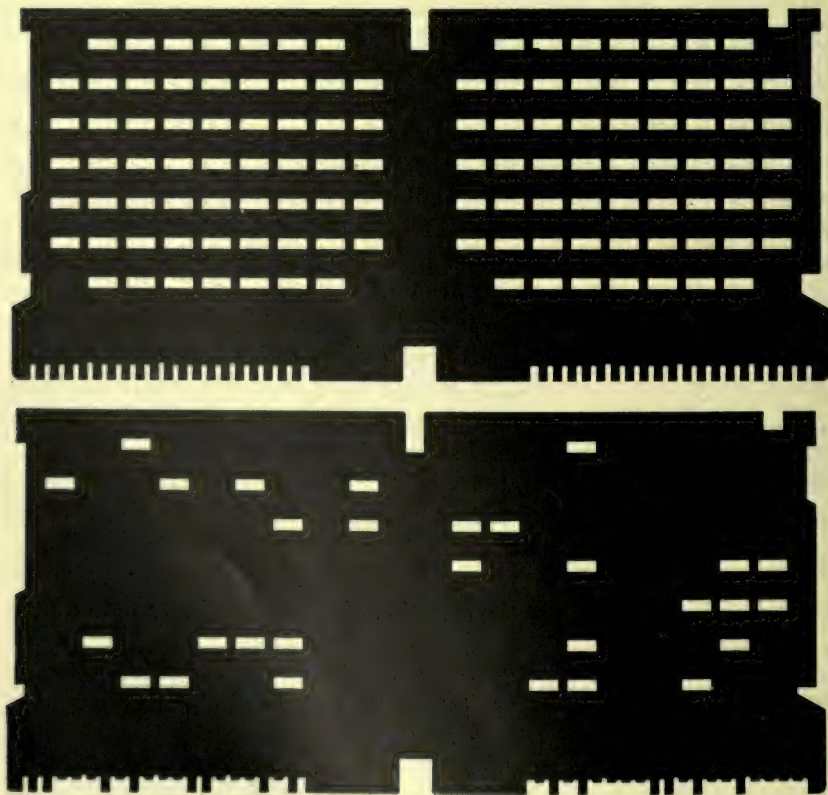


Fig. 5 — Normal (above) and dropped card views as seen from the phototransistor side.

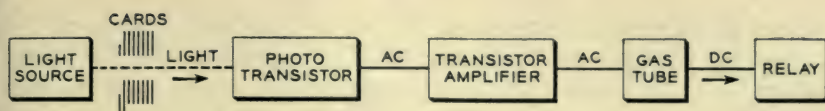


Fig. 6 — Block diagram of channel circuit.

cards in their normal positions they are aligned and the holes in the cards form unobstructed, horizontal tunnels called channels through the entire stack. When a particular card has been selected and dropped a distance slightly greater than the height of an unenlarged hole, these light beams through all of the channels are blocked by the dropped card except for those holes which have been enlarged. This results in a pattern of clear channels which represents the translator output information.

The change in the silhouette of the stack from the condition of all cards normal to that of one card dropped is shown in Fig. 5.

THE CHANNEL CIRCUIT

Fig. 6 is a block diagram of the circuit used to determine whether a particular channel is interrupted by a dropped card or not. Each block, with the exception of the light source, represents a piece of equipment provided individually for each channel. The light source is common to all channels. If the hole in the dropped card for a particular channel has been enlarged, the light will pass completely through the stack of cards and fall on the phototransistor. The phototransistor converts the light into an electrical signal which, after being increased by the transistor amplifier, is used to trigger a cold cathode gas tube. The gas tube in turn operates the channel relay. This relay is located in the associated equipment which uses the information supplied by the translator to process the call.

In making a detailed examination of the channel circuit, it is convenient to consider it in two parts: the optical section and the electrical. The optical section includes everything up to the point where the light falls upon the germanium of the phototransistor. This part of the channel is shown functionally as Fig. 7. The light source is a standard projection type lamp normally rated at 500 watts. To obtain long life it is operated in the translator at about half of its rated voltage at which level its input is approximately 170 watts. This type of lamp was chosen because of its high concentration of light in a small plane area.

The light from the lamp passes through a motor driven perforated disc which modulates it with an approximate square wave at a 400-cycle rate. Modulated light is used because it is more economical to use ac

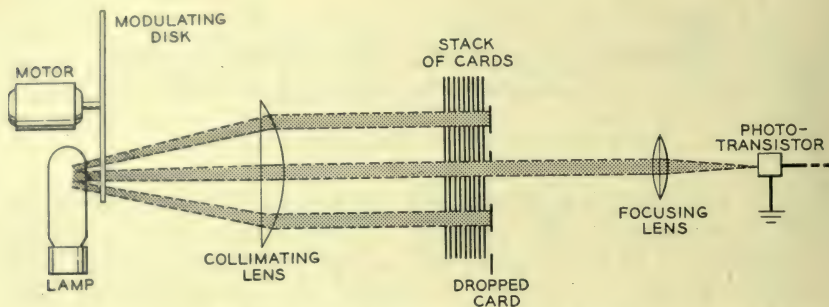


Fig. 7 — Optical section of channel circuit.

than dc amplifiers and also since the difference between the light and dark currents of the phototransistor is the important factor rather than the absolute value of either. The modulated light is collimated by a lens to minimize the loss as the beam passes through the holes in the card. Unless interrupted by a dropped card, this light beam will pass through all of the cards and fall on the lens which focuses the light on the sensitive area of the phototransistor. The light intensity at the lens of the phototransistor is 34-foot candles minimum. This is equivalent to about 12 millilumens at the phototransistor, a figure relatively small when compared to the light intensity that is required by conventional photoelectric cells.

The electrical part of the channel circuits starts with the phototransistor and is shown in Fig. 8. The light acts as the emitter of the phototransistor. The collector is of the conventional type for point contact transistors. As is normal in grounded base transistor circuits, the collector of the phototransistor is biased in the high impedance direction. A variation in the light intensity causes a variation in the collector impedance of the phototransistor. The type used has an impedance of about 10,000 ohms when dark which is reduced to approximately 3,000 ohms when illuminated. The output of an illuminated phototransistor when coupled to the amplifier ranges from 1.3 to 12 volts positive peak at 400 cycles depending upon the age and condition of the transistor.

Since the discrimination by the channel circuit between a clear or blocked light path depends upon the presence or absence of an ac output from the phototransistor, noise of sufficient magnitude, if present when the channel is dark, would cause a false indication. To guard against such false indications each phototransistor is checked during manufacture for dark noise. During a five minute interval the dark voltage must not exceed 75 millivolts.

The phototransistor is coupled to the amplifying transistor by transformer T1. This permits convenient matching of impedances and separation of the dc bias voltage. A voltage limiting varistor, V, is connected across the input of the transformer to limit surges which might otherwise damage the amplifying transistor. The circuit of the transistor amplifier is a conventional arrangement.

Voltage gain of the amplifier, including the input transformer to the gas tube, varies from 40 to 100. However, when operating in the translator, the phototransistors normally will drive the amplifier to saturation

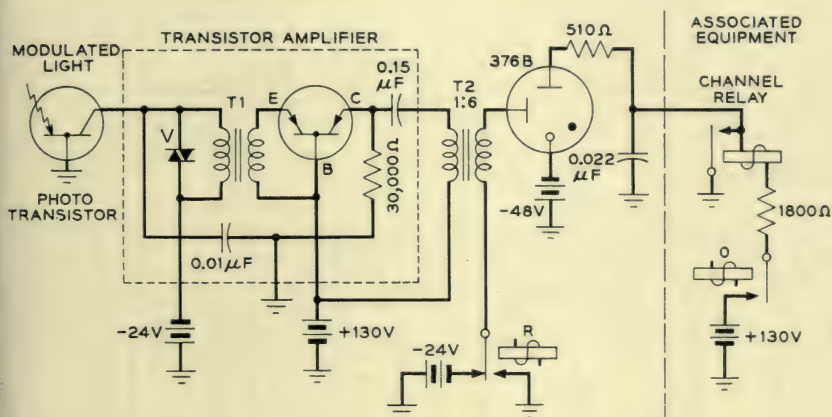


Fig. 8 — Electrical section of channel circuit.

which limits the output to 160 volts positive peak. For the purpose of guaranteeing operation, a minimum output voltage of 38.5 has been set as a rejection point for a phototransistor-amplifier combination.

The output of the transistor amplifier is normally sufficient to break down the control gap of the cold cathode gas tube. Sufficient current flows in this control gap to insure reliable transfer to the main gap when the output control relay O in the associated equipment operates. To aid deionization of the gas tubes, the bias of -24 volts is removed from the control anode just before channel operation is required. The relay R, which replaces the bias voltage with ground, is operated by a circuit which checks that the card being dropped is completely down. This "down" check circuit utilizes the two index holes in the card which are never enlarged and employs two phototransistors to detect the presence or absence of light through these holes.

Fig. 9 shows the card down check circuit. It differs from the routing information channel circuits in that the operation of a relay is required

when the light is blocked. Therefore the transistor amplifier-gas tube circuit is not applicable. For the down check channels, the output of the phototransistors is amplified by the first section of a conventional double triode thermionic emission tube V1. The ac signal at the plate of V1 is rectified by a conventional full wave rectifier consisting of transformer T3 and tube V2. The rectifier voltage is negative with respect to ground and when it is impressed on the grid of the second section of tube V1, that section is driven beyond cutoff. Therefore, as long as light falls on the phototransistor, no current flows through the relay. When the

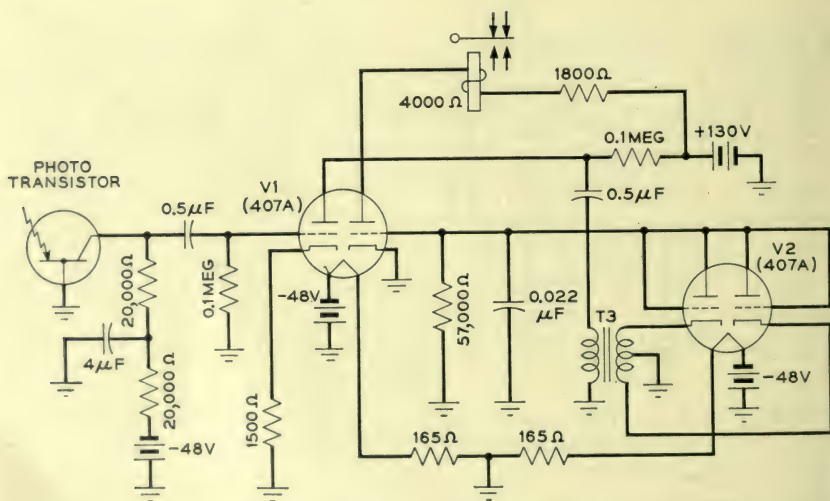


Fig. 9 — Card down check circuit.

dropped card blocks the light, the negative voltage on the grid disappears and the tube conducts, thus operating the associated relay. While the prime purpose of this down check is to signal the associated equipment that the card is in position for recording its output code, advantage is taken of this circuit's ability to distinguish between a light and dark phototransistor at any time to provide alarms in the event of a lamp or modulating disc failure.

After the card has been checked down and the associated equipment is ready to accept the output of the card, that equipment connects a positive 130-volt battery through its channel relays to the main anodes of the gas tubes. All gas tubes associated with illuminated channels will have their control gaps fired and will transfer to the main gaps. This operates the corresponding channel relays in the associated equip-

ment. The relays in operating lock to ground and thereby extinguish the main gap discharges thus increasing the life of the gas tubes. Those channels which have been blanked out will not have the control gaps of the gas tubes broken down. Therefore when the 130 volts is applied, the relays associated with the darkened channels will not operate. The operation or non-operation of the relays in the associated equipment completes the function of the channel circuits.

The capacitor and resistor network at the main anode of the tube is to prevent transients due to the operation of other channels from falsely firing the main gap of a dark channel.

CHANNEL PACKAGES

The phototransistor is mounted in a metal tube along with a lens that focuses the collimated light on the transistor. Fig. 10 is a cutaway view

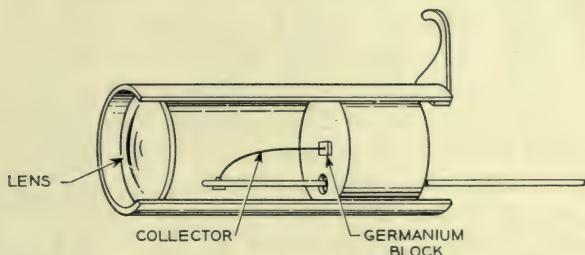


Fig. 10 — Cut away view of phototransistor.

view of the 3A phototransistor showing the relationship of the lens to the transistor. To mount in the translator, the tube is slipped into an accurately positioned hole and clamped in place using the slotted ear. This mechanical fastening is also the ground connection. The output lead from the collector is attached using a slip-on connector.

The amplifying transistor, transformer T1, varistor V, the resistor and two capacitors of the amplifier are packaged in a convenient "plug-in" unit. Fig. 11 is a photograph of these amplifiers. As shown, the transistor is mounted under a removable cap on the package so that it may be conveniently replaced if necessary.

The gas tube, transformer T2, and the associated resistor and capacitor are also assembled as a packaged unit.

TRANSLATOR CIRCUITRY

The card translator is mounted on an associated translator table as shown in Fig. 12. The translator table contains the transistor amplifiers

and the cold cathode tubes one for each of the channel circuits as previously discussed, and, in addition, the miscellaneous relays through which the operation of the translator is controlled. As already stated, the connection between a translator and the associated decoder is through suitable connector relays which are either a part of the decoder or on a separate connector frame.

The phototransistors may be adversely affected by temperatures greater than 130°F. Therefore, to provide satisfactory operation during sustained heavy traffic load periods and with high ambient room temperatures, an air circulating and filtering unit, as shown in Fig. 13, may be provided. This unit mounts on the translator in place of the regular end cover and otherwise requires no further apparatus change. For convenience in ordering, the "1B" translator is specified when the air cooling unit is desired. Otherwise the "1A" translator will be furnished.

The translator table contains relays for controlling the dropping, checking, and restoring of the cards in the translator. First are the relays that operate the card "pull-up" and "pull-down" magnets. Then there are select code bar control relays, one for each bar, operated from the sender or decoder, which in turn operate the associated select code bar solenoids. These relays are necessary since the lead resistance from the sender to the translator would adversely affect the operating time of the code bar solenoids. Finally, there are relays that check the code bars for proper operation on a "two-out-of-five" basis. Two, and only two code

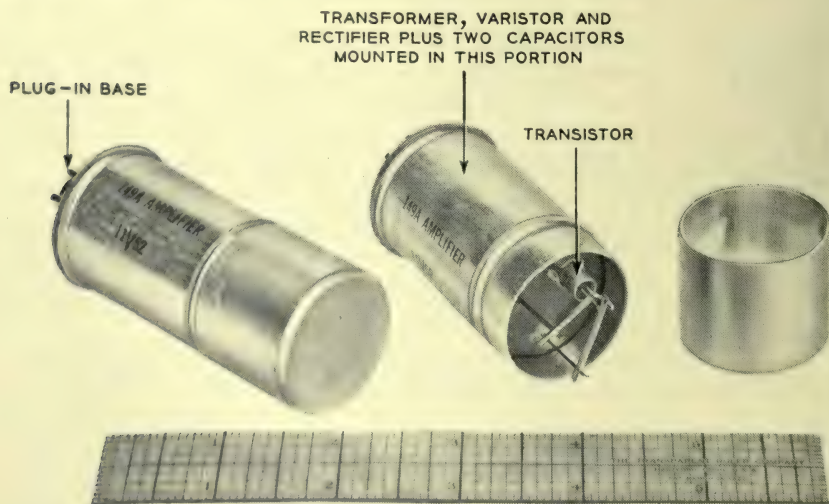


Fig. 11 Transistor amplifier.

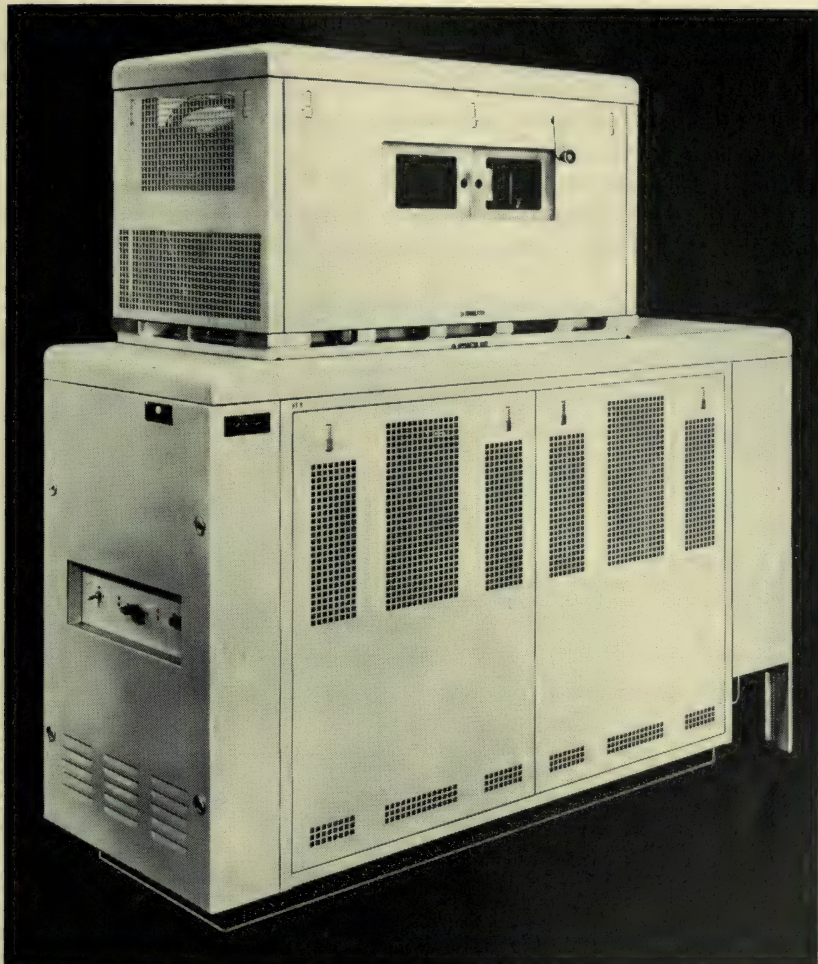


Fig. 12 — Card translator and table.

bars for each digit must be operated, before an attempt is made to drop a card.

The operating cycle of the relays and the translator in selecting a card and restoring it is then as follows:

When the translator is first seized the pull-up magnets are energized and when the cards are suspended the latches are operated. The decoder then closes the leads over which the code bar relays operate and they in turn operate the code bar solenoids. When the proper number of these are operated a check for this is made which releases the latch.

With the latch restored to normal the pull-up magnet is de-energized. All the cards then drop until they meet the code bars, about 0.016". The card for the particular code, however, continues 0.180" further, because the bars for all the tabs have been lowered until it rests on the pull-down magnetic pole face. The index channel relays then operate causing the decoder to read the output of the card.

When the decoder has checked that it has received proper information from the card, and on certain calls when the marker has selected an

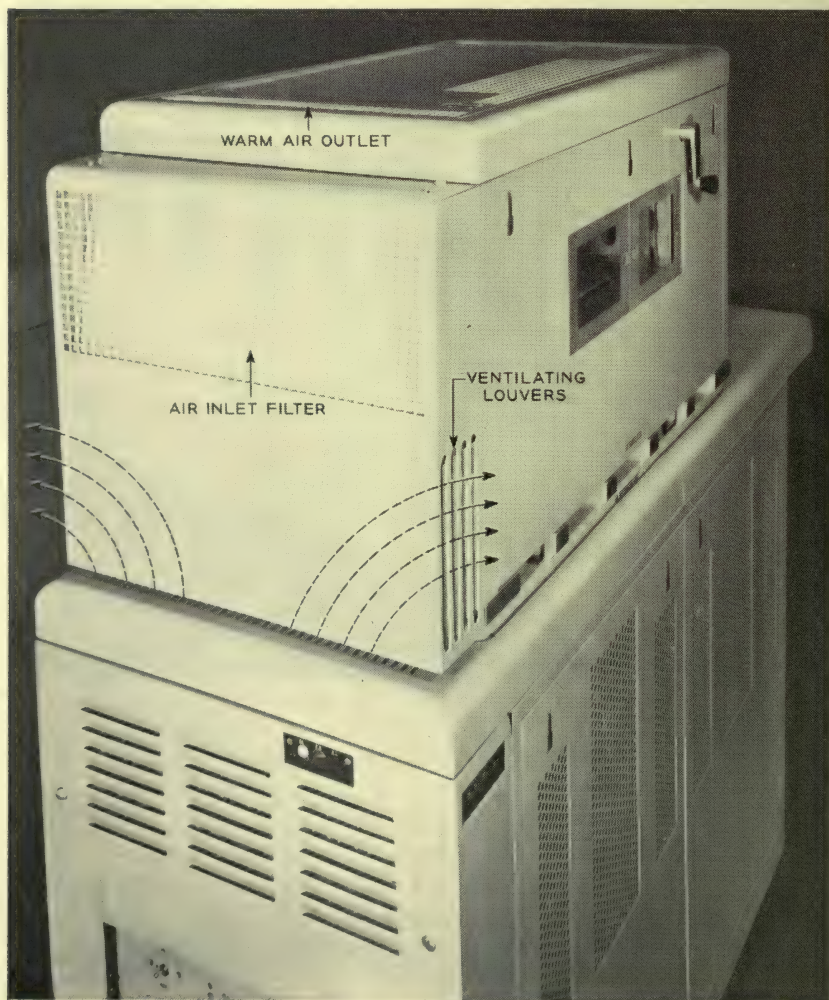


Fig. 13 — Card translator equipped with cooling unit.

outgoing trunk, the decoder releases the translator. Just prior to this, the decoder operates an automatic restoral relay in the translator, which causes the card to be restored to its normal position in the stack. This takes place by again energizing the pull-up magnet, and, when the cards are again suspended, the latches are operated and the code bars are released, restoring to their normal positions. The card that was down is restored by the code bars to its position in the stack. All the relays in the translator table then release except for a slow releasing relay which remains operated holding the pull-up and latch magnets energized so that a subsequent call does not have to wait for these magnets to operate and suspend the cards. This feature saves time under heavy traffic.

TRANSLATION TIME

The call carrying capacity of a translator and the number of translators for a particular CSP is directly related to the time required for translation, that is, the selection, reading and restoral of the translator cards. For this reason, particular attention to translation time was used in the design, not only of the translator, but throughout all the associated circuits of the switching system.

Since the translators in service are always controlled by the decoder circuits and since the number of decoder translators and decoder foreign area (DFA) translators are directly proportional to the number of decoders, the times for the various types of translation and alternate routing will be given in terms of the holding time of the decoder.

PRETRANSLATION

This requires approximately 235 milliseconds. On this type of call translation of three digits, when either four, five or six digits are needed for the routing, informs the sender to release the decoder and wait for the rest of the code.

THREE DIGIT TRANSLATION

The time for this type of call is approximately 330 milliseconds. This is when three digits are sufficient for a routing.

SIX DIGIT TRANSLATION

The time for this is approximately 550 milliseconds, assuming that the six-digit card is in the decoder translator. This does not include the

pretranslation time. Pretranslation does not occur if the office code as well as the area code is registered in the sender before the decoder is connected. This often occurs during periods of heavy traffic. On six-digit translation, two cards must be translated, the three-digit area code followed by the six-digit destination code card.

SIX DIGIT TRANSLATION IN FOREIGN AREA TRANSLATOR (FAT)

This type of call requires approximately 560 milliseconds. The area code card in the decoder translator gives the information for selecting the particular FAT in which the six-digit card is located. The translation time will be extended if there is a delay, due to another decoder using the FAT. Pretranslation time, as stated above is not included.

SIX-DIGIT TRANSLATION-PRINCIPAL CITY AND VACANT CODE ROUTING (FAT)

In this case translation requires approximately 615 milliseconds. This is for routings to areas where there is a principal city (PC), usually another CSP, through which all calls to that area can be completed, although in the area there are other destinations reached over direct high usage trunks. In this case, to save cards and perhaps translators, the six-digit cards for all destinations reached directly through the PC are deliberately omitted. The time given is for a call to such a destination. The three-digit card for the area has on it information for the proper routing of the call to the principal city. For those destinations where the six-digit cards are omitted, as well as for vacant codes in such areas, the call is routed to the principal city. Pretranslation and foreign area translator delay times, if any, are not included.

THREE-DIGIT CARD-TO-CARD TRANSLATION

This type of card-to-card operation is used where there are several sub-groups of trunks or routes to the destination and the decoders do not have facilities for determining which group of trunk has an idle trunk. The routing information for each trunk group must be presented successively to the marker in selecting an idle trunk.

The translation time, considering that an idle trunk is selected from information on the first card, is approximately 330 milliseconds, assuming no marker delays. When the trunks for the first card are found busy and routing is made from the second card, the total time is about 550 milliseconds. This increases to about 770 milliseconds for routings

from the third card and finally to 1060 milliseconds for the fourth card routing.

SIX-DIGIT CARD-TO-RELAY TRANSLATION

This type of operation is used where the first group of trunks to the destination is of the type that requires test by the marker for selecting an idle trunk. There are alternate routes, however, in which the decoder can determine which group has an idle trunk. In this case the area code card provides information for the selection of the first six-digit card. This card has information for routing the call over the first group of trunks. There are alternate route cards for each subgroup of alternate routes. There may be as many as five alternate routes, each of which may have as many as four sub-groups of 40 trunks each.

The translation time, assuming the routing is from the first card, is about 550 milliseconds. For routings from any one of the alternate route cards, the time increases to approximately 800 milliseconds. Pretranslation and FAT delay time, if any, are not included.

THREE-DIGIT CARD-TO-RELAY TRANSLATION

The time required is about 350 milliseconds for routings from the first card and about 580 milliseconds for routing from an alternate route card.

SIX-DIGIT RELAY-TO-RELAY TRANSLATION

This type of operation is used where all of the trunks including the alternate routes are tested by the decoder in determining in which group there is an idle trunk.

The translation time, assuming the routing is from the first six-digit card, is approximately 575 milliseconds. For routings to succeeding alternate routes the time is approximately 800 milliseconds.

MAINTENANCE FACILITIES

Although the card translator and all of its components are designed for relatively long and trouble-free life, adequate testing facilities and maintenance procedures are essential because of the importance of the individual CSP's in nationwide dialing. Adequate guards and methods of procedure have been made available in case of almost any catastrophe that would incapacitate any CSP office. Moreover the complete breakdown of a translator or even several of them in a CSP office would not completely stop calls through that office although the call carrying



Fig. 14 — A trouble record card is being removed from the perforator and information from a punched card is being entered in the office records.

capacity of the CSP office or the switching system would be reduced. However, to insure satisfactory operation of the card translators automatic trouble recording, test circuits and maintenance methods are provided.

TROUBLE RECORDER

In CSP offices, where the card translators are used, the card punching type of trouble recorder will be provided as shown in Fig. 14. In the event of a failure of any translator, the associated decoder circuit will block and time out. This causes the connection of a multiplicity of leads between the trouble recorder and the decoder as well as to the translator. The trouble recorder will then automatically punch and drop a properly designated paper card (Fig. 15) that shows which translator and decoder are in trouble. The state of the various elements of the translator and decoder will also be recorded. In addition, a record is made showing which code bars, pull up and pull down magnets, latches and important relays are operated. Also the important key relays which are operated in the decoder will be shown. The associated sender will be identified as will also the marker if one is connected. Then too, the state of the marker will be recorded. From this trouble record the cause of the failure can, in most cases, be determined.

When the trouble recorder card has been punched, the associated

equipment is directed to make another trial. This usually will be with an alternate decoder and translator so that the call is usually completed with a delay of a little more than one second required for punching the trouble recorder card. The trouble recorder, once it has completed punching a card, is immediately available for recording another failure. The trouble recorder is also available for recording failures of other CSP equipment such as the controllers, senders and markers in a manner similar to that described for the decoder and translators.

TRANSLATOR TEST

There are facilities provided on the trouble recorder frame, through keys and connecting relays, for adding and removing cards from the translators. On adding a card, a check can be made to verify the card output to make sure that it is in agreement with the template from which the card was coded. Test calls using all decoders and markers can be made to verify the selection of a trunk in accordance with the routing information on the new card as well as on any card in any translator. These tests can be caused to recycle automatically and to drop a trouble recorder card in case of failure. This feature is useful in isolating infrequent failures in any of the associated translators, decoder or markers. Facilities are also provided for removing the translator selector unit for periodic inspection and adjustment as may be necessary. Also timing tests can be made from the trouble recorder frame to check the time required for the translator to drop and restore a card.

To determine if any of the output channel elements of the translator, which includes the photo-transistors, transistor amplifiers and cold cathode tubes, for all the channels, have satisfactory operating margin, means are provided for making a mass test of the channels. This test is made under a controlled voltage (36.5 volts) which is considerably below the worst service condition. In case of a failure of an element under this test, a trouble record card will be dropped which will show, by a punched hole, each element that is operating satisfactorily. This mass test is provided to detect any channel element that is approaching end of life and thereby assures that there is at all times ample operating margins of these important channel elements.

Portable Test Set

In addition to the test facilities provided on the trouble recorder frame, there is also a portable test set for the translators as shown in Fig. 16. This test set is connected to the translators by means of multicontact

2-3845 (1-52)

	0	5	10	15	20	25	29	
S8	SOURCE OF RECORD				DECODER			
	D	M	C	DT	MT	TV	CT	
S7	TYPE OF RECORD				MARKER-TC			
	FIF	MFT	CFR	DST	MTST	RTRF	TST	
S6	SDR, FR TENS				SDR, FR UNITS			
	RO	PRO						
S5	TRANSLATOR ENGAGED				DECODER CONNECTOR			
	H	EM	TO	T1	U0			
S4	DECODER INPUT CODE				CODE BARS			
	A0	1	2	4	A7	B0	1	
S3	DECODER INPUT				LATCH MAGNETS			
	A0	1	2	4	A7	B0	1	
S2	DECODER ROUTE ADVANCE				CARD GROUP			
	3D	6D	6DA	VO	NVO			
S1	RA RAI RA2 RA3 GSO GSI GS2 GS3 GS4 GS5 GO GI G2 G3 GLB RLS MB RO ROIT ROTC				MRR ROUTE ADVANCE			
	NCA	CA4	CA5	CA6	IT	TC	ITC	
R8	ALTERNATE ROUTE				ROUTING INST.			
	COLCINAC	AC	AHA	AF	TO	1	2	
R7	NSK SK3 SK6				CODE CONVERSION			
	DIM	RD						
R6	TRUNK BLOCK CONNECTOR				GROUP START			
	TO	1	2	U0	1	2	4	
R5	DECODER ROUTING INSTRUCTIONS				MARKER REGISTRATION-CODE CONVERSION			
	CC	CR	RR	FOF	FMB	FRO	FST	
R4	MARKER REGISTRATION				MRR-SDR TRANSMITTED CODE CONVERSION			
	MB	RO	PRO	FOF	FMB	FRO	FST	
R3	TRANSMITTED-MARKER TO SENDER				FRAME IDENTIFICATION			
	M	DC	MF	SXD	LPD	XSD	DLC	
R2	DECODER PROGRESS				MARKER PROGRESS			
	CKG	HTK	SMI	SMC	CK3	CKK	CKH	
R1	MARKER PROGRESS				MRR-SDR TRANSMITTED			
	DCB	DCB2	ATB	GPL	ARS	TCD	TKS	
R0	MARKER PROGRESS				MRR-SDR TRANSMITTED			
	AK	BK	CK	CHS	A	C	HMG	

IBM 907501

Fig. 15 — Trouble recorder card showing the multip

plugs, cords and jacks. This set can be used for adding and removing cards and verifying that the new card will drop. Bins are provided for storing cards in the process of adding, removing, and transferring cards. Timing tests can be made not only of the over-all time of the translator, as from the trouble recorder frame, but also of the individual component parts of the translator. The test set can also be used for removing the translator selector unit and for bench testing this selector. In this connection, a re-cycling test can be made of the code bars, card support bars and latches to check that they operate smoothly and evenly. Current flow adjustments can be made of the code bar solenoids, latches, and pull up and pull down magnets.

DESIGN OBJECTIVES

When, in the development of the development, the controlling requirements that had to be met became apparent, the design objectives were considered. The reasoning applied is set forth in retrospect in the following paragraphs.

It will be apparent from the foregoing that administrative problems would be involved should it be necessary to arrange and maintain the individual cards in any particular order in the card stack and, therefore, indiscriminate loading became a design objective.

Since the cards will be changed from day-to-day and sometimes on

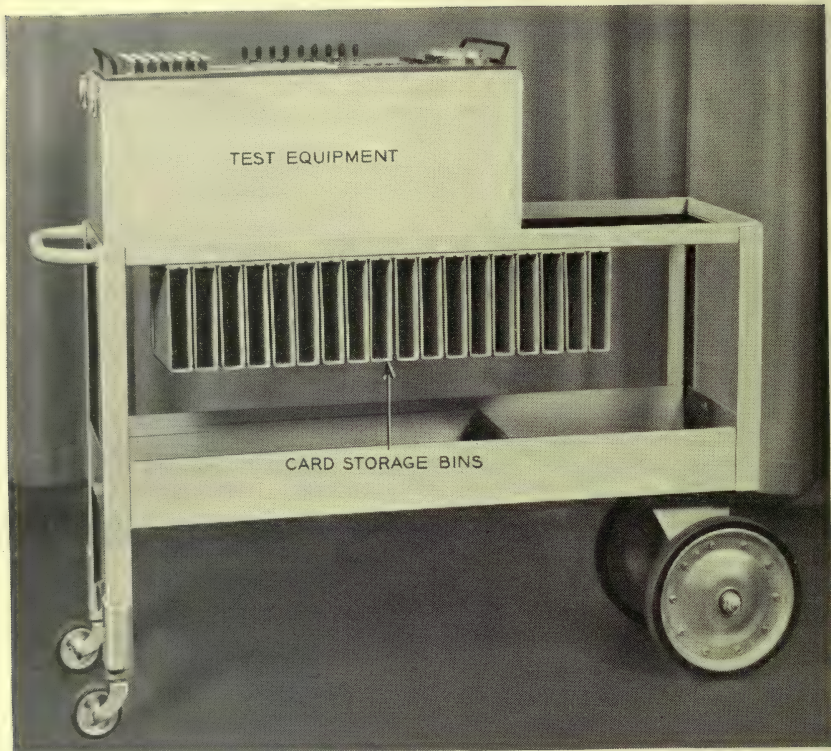


Fig. 16 — Portable test set.

It was estimated that a card translator, on the average, would be called upon to function from twenty to thirty million times a year. Accordingly, provision for direct-acting, high-speed, long-life and readily replaceable components became another objective.

A further objective was the provision of means for reading the routing information without mechanically contacting the cards, as for instance by utilizing photo-detection circuits, such as have been referred to, as it was reasoned that in this way reliability over an extended period could best be assured.

ELECTRO-MECHANICAL DESIGN

Phototransistors are utilized as the photo-sensitive elements, the light source is a standard projection lamp, the light beams are modulated and then directed through the card formed tunnels by dual collimating lenses and then upon emission from the card stack are focused on the phototransistors. (See Figs. 17 and 18.) To a large extent the card, which is

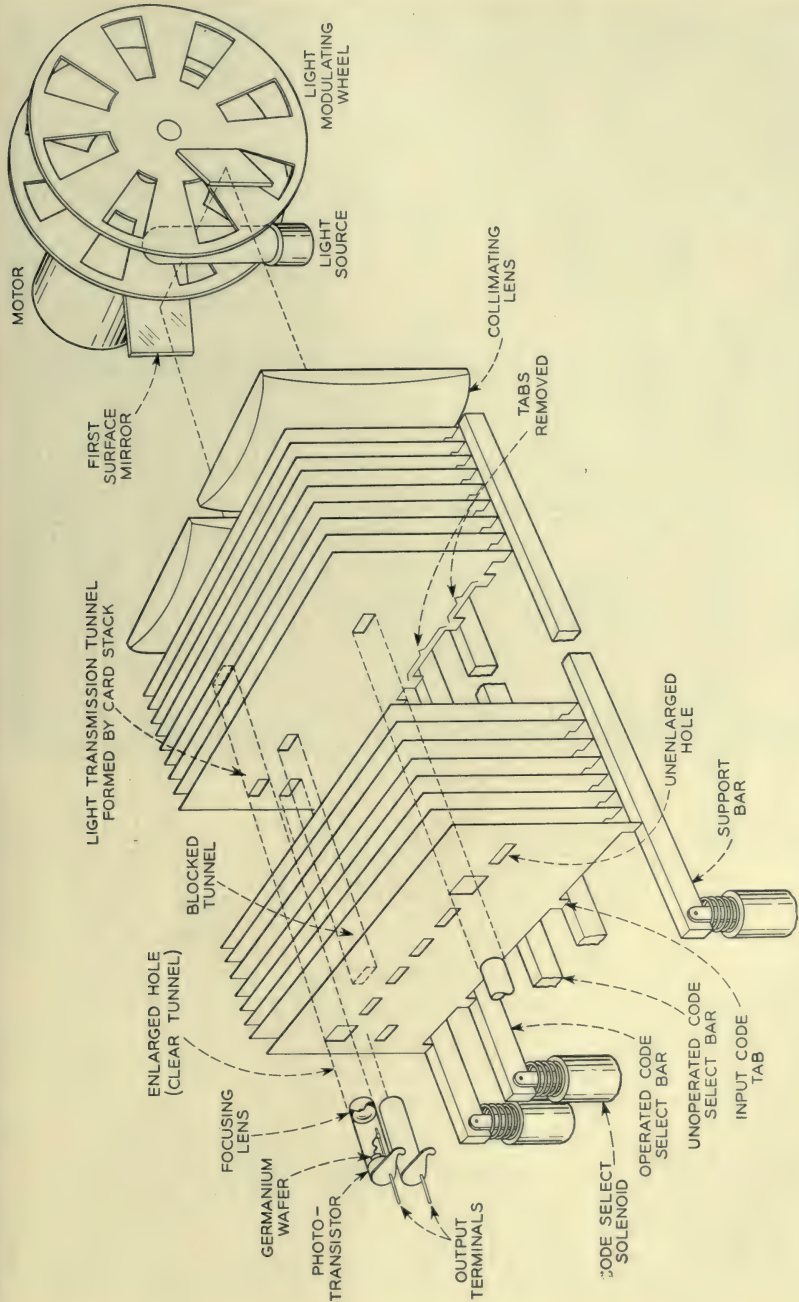


Fig. 17 — Functional mechanical schematic.

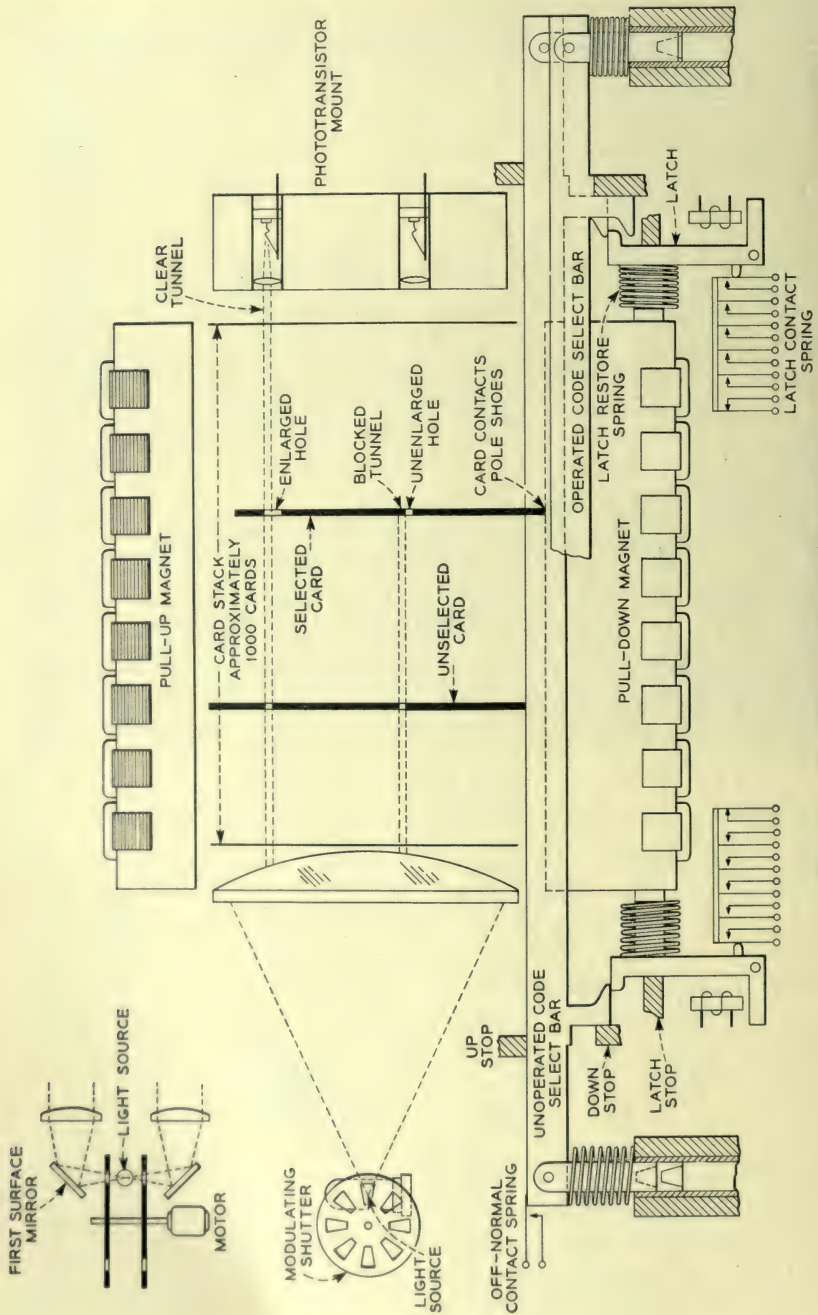


Fig. 18 — Simplified mechanical schematic.

having done so, any red can be seen, more holes have been enlarged and/or more tabs have been removed than the template calls for. If, on the other hand, openings in the template are visible through holes in the card that have not been enlarged, the hole coding is not complete. If tab notches appear in the template other than in registration with the portion of the card from which tabs have been removed, the tab coding is not complete. Accordingly, the template also serves as a convenient means for checking the coded card for accuracy. As a convenience, the holes of the template were made round and its tabs triangular.

The holes of the card before being enlarged are of a form and size that simulates the filament face area of the light source. To accommodate the 118 holes, the optimum dimensions of the holes were determined to be $\frac{3}{8}$ " wide, 0.140" high and the optimum vertical and horizontal spacing proved to be 0.535" and $\frac{1}{2}$ ", respectively.

The tabs are of a form and size determined largely by the mounting space required for the solenoids that are used to operate the code select and card support bars, the vertical displacement required for shuttering the holes and ruggedness considerations. Thus, the nominal size of the tabs associated with the code select bars became $\frac{1}{8}$ " wide $\times$ 0.205" long with a spacing of $\frac{3}{16}$ ". The tabs associated with the card support bars — one at either end — being used for each translation instead of

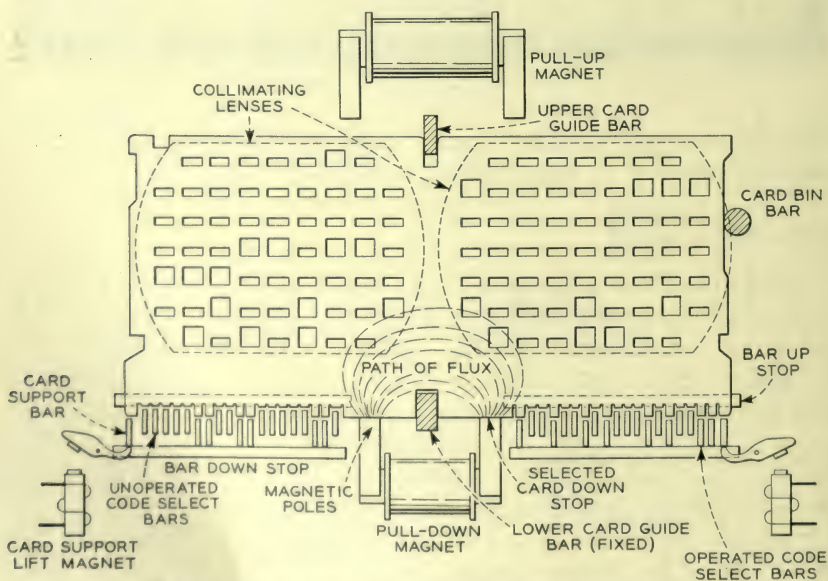


Fig. 20 — Card and directly associated components.

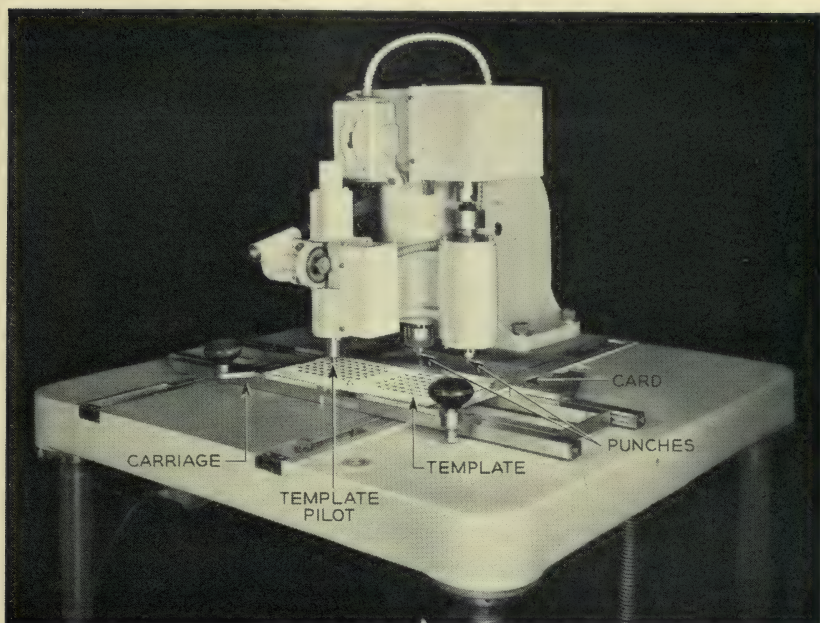


Fig. 21 — Card coding tool.

only selectively as in the case of the tabs associated with the code select bars were made $\frac{1}{4}$ " instead of $\frac{1}{8}$ " wide and their outer corners were rounded generously. The additional ruggedness thus provided is particularly beneficial because the end tabs are subject to more abuse than the others in handling.

Left and right-hand grouping of the holes and tabs was necessary for the dual lense system and to provide space in the central area for guide notches and for mounting the pull-up and pull-down magnets, as shown in Fig. 20.

Center, top and bottom notches are provided for engagement with card guide bars that position the cards for proper alignment of the holes and tabs. In addition, these notches guide the cards during their operative displacement.

Other marginal notches are provided for mass lifting of the cards, end-for-end positioning in the card cage, locating the cards properly in a tool for coding cards, and for another tool which facilitates handling the cards in bulk. The coding and bulk card handling tools are illustrated respectively by Figs. 21 and 22.

The over-all size to accommodate the various elements and to provide adequate strength is $10\frac{3}{4}$ " wide $\times$ 5" high.

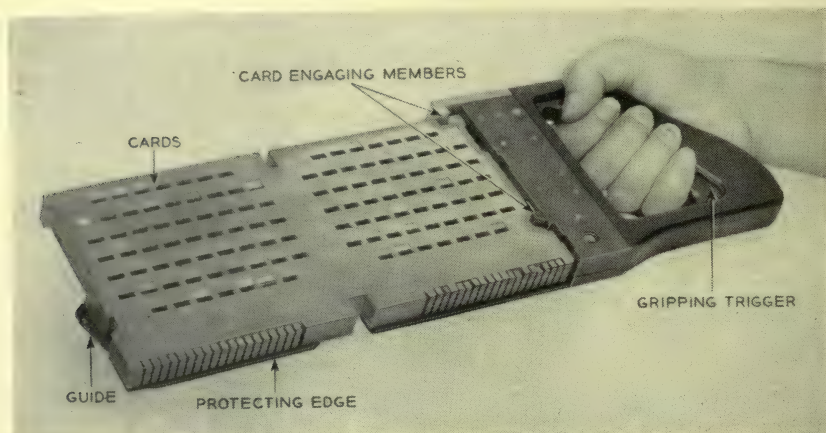


Fig. 22 — Bulk card handling tool. One side plate has been removed to show cards.

The card material had to be magnetic so that electromagnets could be used to assist in the manipulation required in the course of a translation. After giving due consideration to the use of permalloy, Armco iron and several other materials it was concluded that the most practicable compromise would be to use hard-rolled AISI-C1008 strip steel.

The protective finish selected was chromium flash over nickel-plate. The nickel provides reasonable protection against corrosion, while the chromium provides surface characteristics consistent with the need for maintaining interface friction and wear at low levels. The finishes are applied to the strip stock by a continuous plating process and it may be interesting to point out that the current value when applying the chromium is in the order of 1,000 amperes. It will be realized, therefore, that one of the problems in producing satisfactory material is maintenance of the rolls over which the material is fed and which also are circuit elements, free from foreign materials because otherwise high resistance points may develop and cause burning.

Flatness is important because any deviation from a plane, in effect, increases the thickness of the card and thus limits the card capacity of the translator. Manufacture is controlled so that the slight bowing that results from handling the material in rolls always is in the same direction as the material is fed to the perforating and blanking tools. Out-of-flatness is held to maximum 0.012" within 72 hours after fabrication. Bowing subsequently increases the out-of-flatness but not to an important extent. The out-of-flatness is checked by means of the gauge illustrated by Fig. 23, which was developed especially for the purpose.

It comprises a flat metallic plate upon which the card is laid, two rails which are insulated from the plate and project about it an amount equal to the thickness of the stock from which the cards are made plus the allowable amount of out-of-flatness and a metallic roller that is sufficiently long to span the rails. The rails and the flat plate are elements of a neon tube detection circuit which is arranged so that if the roller in passing over the card contacts it the tube will fire. If it does not fire, the card is within the flatness setting of the gauge.

To meet the 0.012" allowance consistently it was necessary to resort to special stress relief annealing. In developing the process it was found that the cards have to be degreased thoroughly and a clean condition has to be maintained at all times. The cards are clamped under heavy pressure between thick nickel plated platens and are then heated and

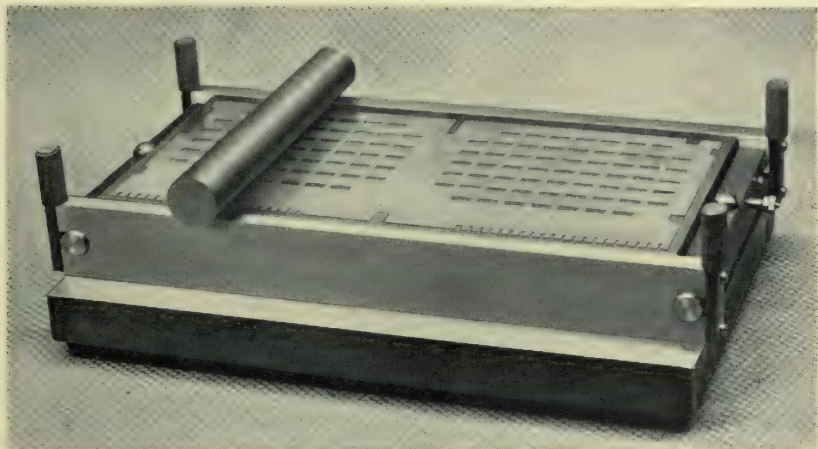


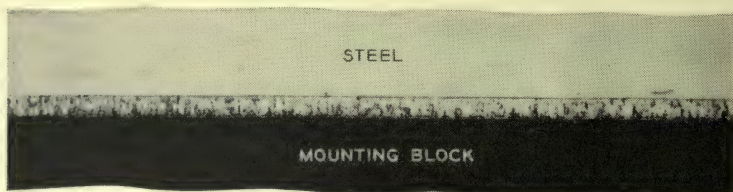
Fig. 23 — Card flatness gauge.

later cooled in a reducing atmosphere. Such a procedure would seem to be reasonably straightforward, but it was found that the heating cycle has to be controlled precisely as otherwise the surface is roughened.

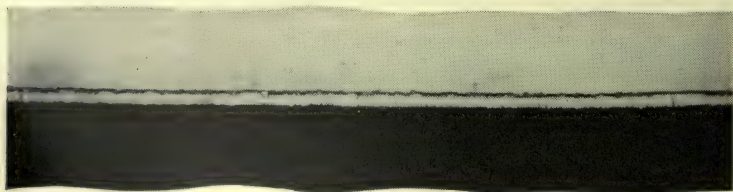
Smoothness of the cards was found to be related to surface color. Improperly stress relief annealed cards usually assume hues of blue and brown which were traced to surface films of so little thickness — perhaps in the order of 500 to 1000 Angstrom units as to cause interference colors. It was found that when the surfaces are discolored excessively, there often is an appreciable increase in interface friction, sometimes amounting to 30 per cent. This is undesirable because free action of the cards is of paramount importance and of course, wear should be minimized. It

was found that roughness is caused by cannibalistic grain growth of the nickel after recrystallization; that is, some of the grains grow at the expense of others, thus causing surface roughness. In this connection it may be interesting to refer to the photomicrographs of Fig. 24. In each, the top, light and almost structureless portion depicts the basic material (steel). The adjacent layer characterizes the nickel plating, the columnar structure being typical of the electro-plating process. The next layer shows the chromium flash and the black portion, a block used to mount the specimen.

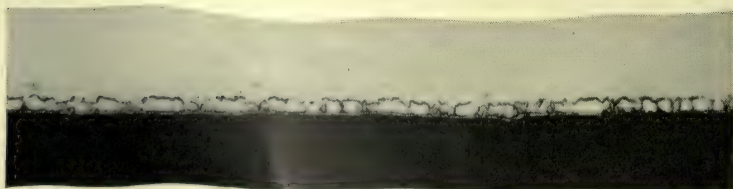
The top view of Fig. 24 is representative of the material as received from the supplier. The center view illustrates satisfactory stress relief annealing. The bottom view illustrates unsatisfactory annealing. It will be noted that after satisfactory annealing some recrystallization of the nickel has taken place and that grain growth has started but has



MATERIAL AS FURNISHED



MATERIAL SATISFACTORILY STRESS-RELIEF ANNEALED



MATERIAL ANNEALED AT TOO HIGH A TEMPERATURE

Fig. 24 — Photomicrographs of sections of cards (1000X).

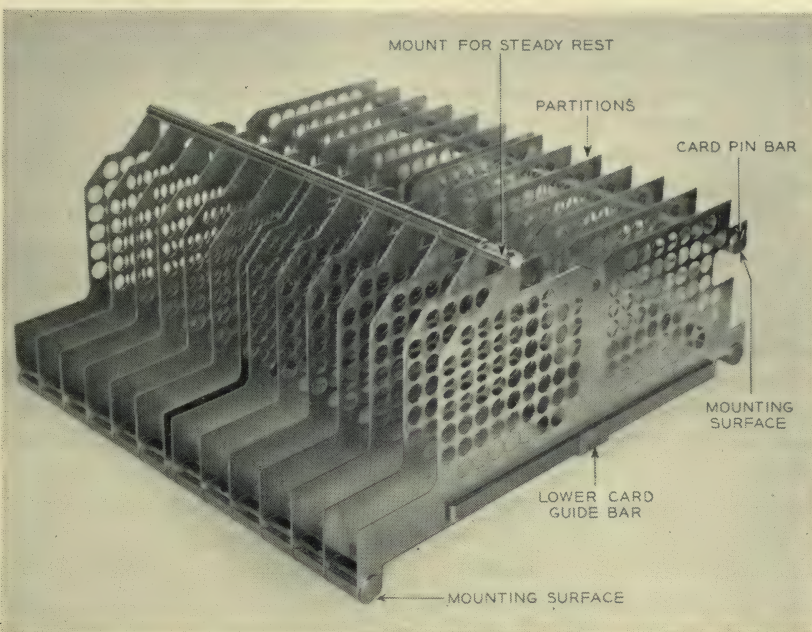


Fig. 25 — Card cage.

not progressed far enough to affect the surface sensibly. It will be observed that when the temperature is too high the grain growth is more pronounced and that as a result substantial protuberances appear on the chromium surface. Cards possessing such protuberances exhibited a marked increase in interface friction.

Distribution of the cards in the translator posed some important problems. Sufficient space had to be provided to accommodate a full complement of cards. Yet it was realized that there would be cases where the translator would not be fully equipped and under these conditions the cards, due to the surplus space available, might lean sufficiently to render them inoperable. To obviate such a possibility a partitioned cage — see Fig. 25 — is provided for the cards. The partitions are close enough to one another to assure that if but a single card is used in a compartment it will operate satisfactorily. This is illustrated by Fig. 26. The spacing thus arrived at is a little less than 1". This spacing provides sufficient room for reliably handling 98 coded cards and a blank card adjacent each partition making a total of 100 cards to a compartment. The blank cards improve the performance of the adjacent cards.

Twelve compartments were decided upon after due consideration of the many related problems such as light tunnel transmission losses,

and code select bar deflection. Twelve compartments fully loaded can accommodate nearly 1200 cards. Thus a comfortable margin beyond the 1000-card capacity requirement has been provided.

Card thickness was controlled by considerations such as stiffness of the tabs, flatness and overall length of the card stack. It ultimately was concluded that the optimum thickness is 0.007" nominal.

The weight of a card after coding averages 40 grams and, considering that nearly 1200 cards may be in translator the total card load is approximately 100 pounds. This is considered to be enormous from a telephone switching apparatus viewpoint and largely accounts for the need of the pull-up magnet.

Manipulation in effecting a translation has to be fast. The displacement of the selected card has to be comparatively large and the card stack, which also has to be moved for each translation but to a lesser extent, is exceptionally heavy. These are not compatible conditions but by advantageous coordination of the pull-up magnet, the pull-down magnet, the latches and the card support bar lift magnets, it has been possible to meet them reliably.

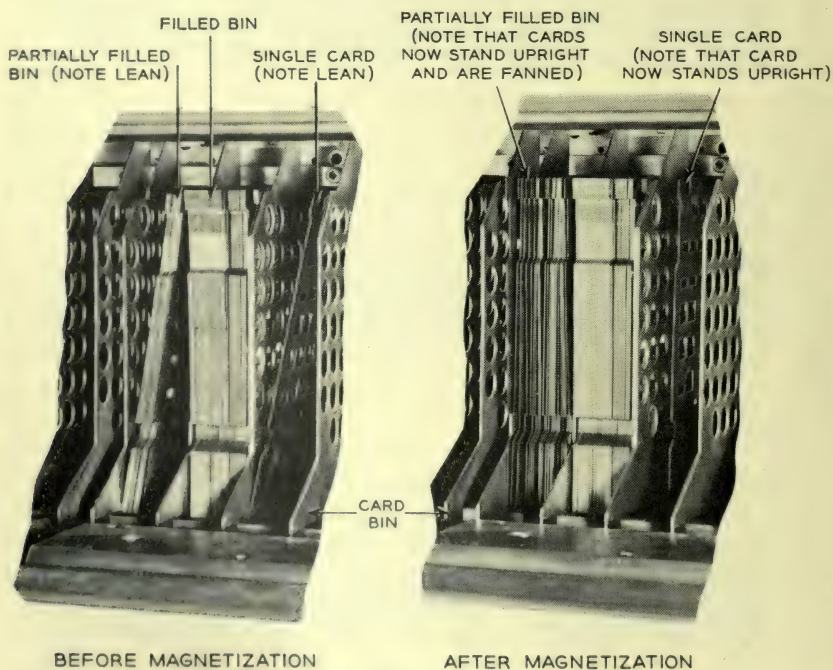


Fig. 26 — Card stack before and after magnetization.

The weight of the stack normally is borne by the latches via the code select bars, but as the time approaches for their operation the stack is elevated by energization of the pull-up magnet and the latches thereby are freed for fast operation.

Separation of the cards just prior to dropping a selected card is effected by the pull-up and pull-down magnets. These magnets induce like magnetic poles in corresponding portions of all the cards and the partitions of the card cage as may be deduced by reference to Fig. 20 and the resulting mutual repulsion causes the cards to separate as may be noted by reference to Fig. 26. This assures free action of the selected card. The pull-up magnet is deenergized when the required code select bars are checked down. The pull-down magnet is maintained energized to assist the dropping of the card.

The nominal initial jump of the card stack, or the separation between the top of the card stack and the pull-up magnet pole faces is nominally 0.016". When the magnet is energized the stack "jumps" to the pole faces in about 15 milliseconds which is fast considering that the load may be as much as 100 pounds.

The nominal drop of the selected card is 0.180". Accordingly, the theoretical free fall time is 33.2 milliseconds and the terminal velocity is 12.7" per second. However, under working conditions, the selected card does not fall freely and except for the added pull of the pull-down magnet, the drop time would average about 40 milliseconds. The pull-down magnet is capable of reducing the drop time in a working translator to as little as 16 milliseconds or approximately one-half the theoretical free fall time with a measured terminal velocity as high as 35" per second. Should this speed be permitted, the impact forces developed would be of sufficient magnitude to cause the tabs of the cards to mushroom to an unworkable degree and to cut through the nylon surfacing of the pole faces of the pull-down magnet in less than one million operations. Under the restraining influence of the card support bars, the drop time averages 33 milliseconds or a reduction of approximately 20 per cent with respect to the unaided operate time and the maximum terminal velocity measured has been 20" per second. Working at this speed, there is virtually no mushrooming during the normal life of the card and the nylon facing survives well over 100,000,000 operations, thus showing a gain attributable to the use of card support bars of more than 100 to 1.

Inadvertent dropping of all cards of the stack is possible although highly improbable. It can happen if the stack is released from its suspended position when the latches are not in a position to support the code select bars. The combined strength of retractile springs of the

solenoid operated bars is insufficient to support the load and, therefore, if this happens, the bars will yield until the load is taken up by the pole faces of the pull-down magnet. The total drop amounts to 0.180" which then will become the effective air gap at the pull-up magnet pole faces. This is more than ten times the normal gap (0.016") and the pull-up magnet is incapable of restoring the card stack to its normal level. This is despite the fact that when the gap is normal, the pull exerted is in the order of 400 lbs and the breakaway pull is approximately 1000 lbs. Accordingly, mechanical means had to be provided for restoring the stack to normal under this condition. Since this rarely occurs, it was concluded that manual restoration would suffice and a hand crank, together with sundry components, including card lift bails, that are illustrated by Figs. 27 and 28 are provided.

Coding of the cards involves many considerations. It has been mentioned that a template is used to indicate the holes that should be enlarged and the tabs that should be removed. Obviously, punch and die sets are required to do this. They are provided as components of a tool which is illustrated by Fig. 21. This tool provides a carriage upon which the card to be coded and the template that provides the coding information are mounted in fixed relationship. The carriage may be moved about to effect registration of the holes to be enlarged or the tabs to be removed, with respect to the punches and dies. When enlarging holes, the carriage is moved to a position where a foot treadle operated pilot registers with a hole in the template and then the treadle is depressed. The pilot enters the hole in the plate that supports the template. The hole is close fitting to effect approximate alignment. As the motion progresses a pilot that is part of the punch assembly effects precise alignment between the punch and the hole to be enlarged. This final alignment is made possible by providing for a small amount of float of a card on the carriage. As the foot treadle is depressed further, the punch enlarges the hole in the card. For coding the tabs, a second punch and die set is provided. It is brought into action by a microswitch controlled solenoid that actuates the punch. The carriage referred to is notched along one edge in registration with the tab positions of the template. Wherever the template is notched, the corresponding card tab is to be removed and the carriage is positioned so that these notches successively may be registered with a stylus associated with the solenoid control switch. The carriage then is moved against the stylus and in doing so the switch is operated which causes the solenoid to actuate the punch, thereby clipping off the corresponding tab of the card. In clipping off tabs the punch does not have to be aligned precisely with the tab

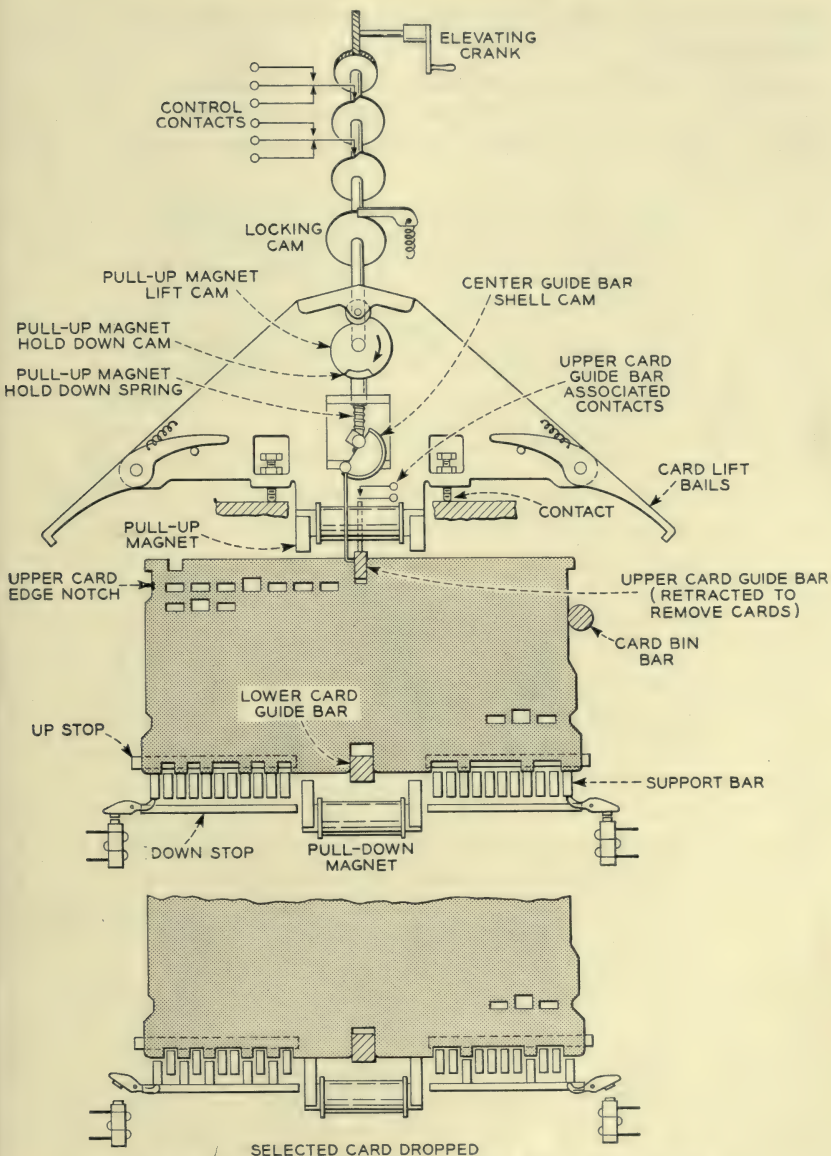


Fig. 27 — Interrelationship of card select components.

position and, therefore, the tab punch has not been provided with a positioning pilot.

Freedom from burrs after coding is assured by passing a sulphur-free abrasive bearing rubber pad over the card several times.

Care is required in handling the cards to avoid distortion of their

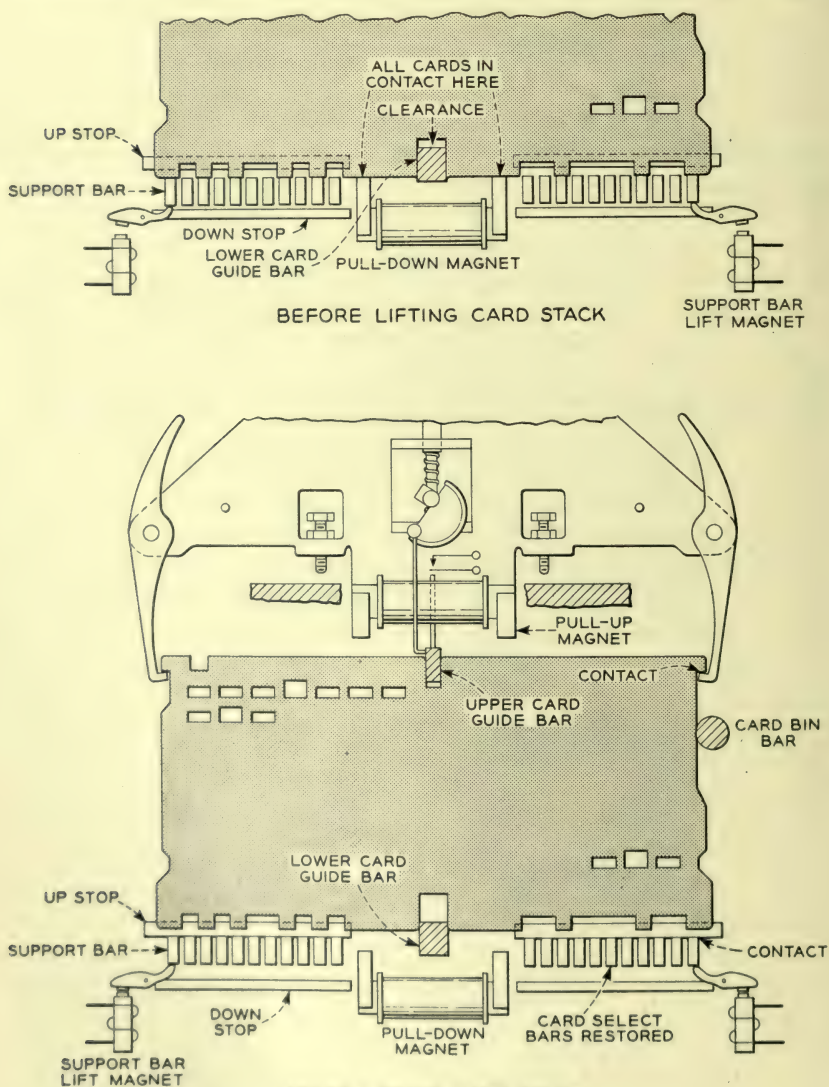


Fig. 28 — Mechanical lifting of cards.

tabs. As a precautionary measure, tools have been made available that afford adequate protection of the cards while they are being handled. One of these tools is for inserting one card at a time in the card stack. The other tool, which is illustrated by Fig. 22, is used when it is necessary to handle a number of cards at a time. It will be noted that it affords protection for the tabs. This tool may be used for loading and unloading all of the cards of a compartment of the card cage in a single operation. The cards, after bulk removal, may be conveniently stored in one of the card compartments of the test set illustrated by Fig. 16.

CODE SELECT AND CARD SUPPORT BARS

The load on a single bar may be as much as 24 pounds. Deflection due to loading has to be maintained at a minimum because any deflection reduces the effective height of the light tunnels through the card stack. However, the bars have to be light because high-speed operation is required. In addition, they have to be wear-resistant because of the many millions of operations to which they are subjected annually. These are difficult conditions to meet but a satisfactory solution was found by providing deep section die cast magnesium bars that are fitted with details made from suitable wear-resistant materials at the point of loading. Such bars, as components, are depicted by Fig. 29, in which the bottom bar is of the card support type. The other three bars shown are code select bars. They illustrate the different terminations that are required by the mounting arrangement employed.

The spacing is $\frac{3}{16}$ " to affect alignment with the tabs of the cards. This close spacing makes it necessary to vary the amount of overhang beyond the up and down stops so that the associated solenoids may be mounted in staggered array and thus their diameter need not be unduly restricted. Despite the variations in conditions it was made possible to use but a single mould by casting the ends of the bars of uniform cross-section and long enough to accommodate the longest extension required. For those cases requiring lesser extension, the bars are cut off to suit. The bars are designed so that they may be used end-for-end and, therefore, although five mounting positions have to be taken care of, the three different end terminations illustrated suffice. It may be noted that the bars during fabrication are straightened if necessary by bumping. In this way, the necessity for side surface machining is obviated and in consequence the side skin of the bars is maintained intact which enhances stiffness.

The weight is only 115 grams each and yet their deflection under the heaviest working load that may be developed is but 0.003" which is

immaterial insofar as reduction in the effective height of the card stack light tunnels is concerned. The stresses developed are so low that fatigue failures should not be experienced.

Nylon is bonded to the top surface of the bars because of its proven ability to resist the cutting action of the cards. Details made from graphitized phenolic linen are secured to the bars at the lateral guide points because of the superior ability of this material to resist wear under the sliding action involved. This material also is well adapted to resist the heavy loading that is developed where (top of center slot) the card support bars are engaged by the tongues of the card support bar lift magnets and it is, therefore, used here also. Beryllium copper details are affixed to the ends of the bars where they are coupled to the associated solenoids. Beryllium copper is used here because of its wear-resistant and non-magnetic characteristics. For the same reasons it is used for details that are secured to the bottom edge of the bars where end guides are engaged and the bars cooperate with the latches.

The beryllium copper details affixed to the bottom of the code select bars are each provided with toe-like projections. The bottom of the toes

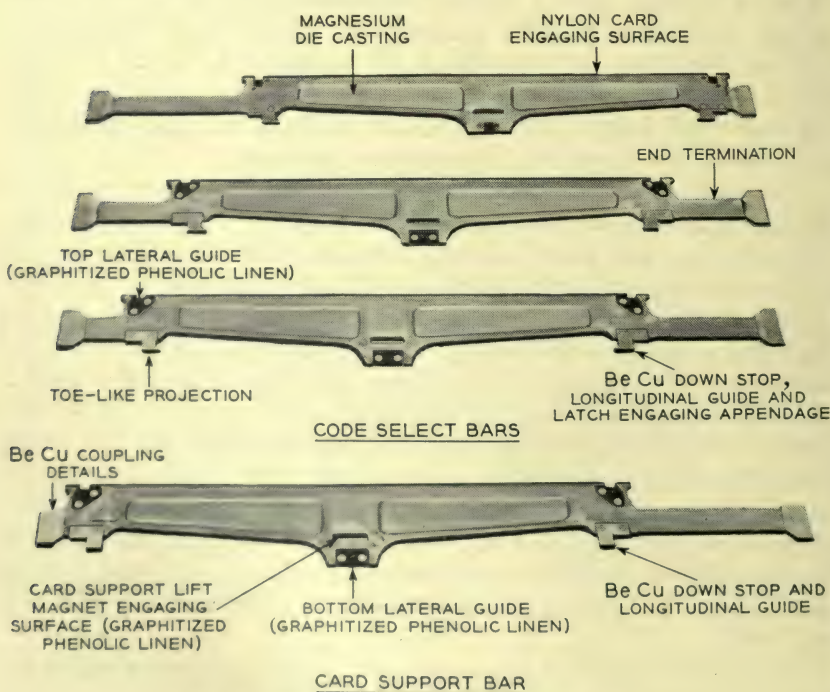


Fig. 29 — Code select and card support bars.

normally rests on the support blade of the latches. After the pull-up magnet is energized, support by the latches no longer is required and they are operated to move their support blades out from under the toes of the code select bars, so that the combination of bars called for by the code digits may be depressed. After the combination is checked down, the latches return to normal so that they are again in position to support the card stack. The toes of the depressed bars project under the support bars of the latches, thus tending to prevent any of the depressed bars from restoring prematurely. The card support bars are not provided with toe-like projections and therefore, may be depressed regardless of the position of the latch support blades. This is necessary because the card support bars have to operate after the depressed code select bar combination has been checked down and the latches restored. The selected card then rides the support bars down and controls the terminal card velocity within safe limits.

Transverse alignment of the code select and card support bars within close limits is important to proper coordination of the bars with the tabs of the cards and also to maintain transverse operating clearance between the bar coupling details and the blade like extension of the plungers of the associated solenoids. It also is necessary to control the up-stop level of the bars accurately so that the close vertical alignment of the holes in the cards that is required may be assured. These conditions are met by the provision of details that serve both as transverse guides and as up-stops for the bars in groups of twenty each which is consistent with their left and right-hand grouping. Since these combined stop and guide details are needed at both ends of each group, four are provided. They comprise details that are stepped to provide a horizontally disposed surface that serves as a common up-stop and a vertically disposed wall that is slotted to form a comblike transverse guide for the solenoid operated bars. The nylon surfaced top of the code select and card support bars engages the bottom surface of the combined up-stop and comb guide details under the action of the solenoid retractile springs. The comb portion engages the graphitized phenolic linen details that are set into the code select and card support bars near their top at both ends.

The lowermost position of the solenoid operated bars has to be controlled in order to establish their proper operated position with respect to the pole faces of the pulldown magnet and the bars have to be guided endwise so as to maintain a longitudinal operating clearance at the coupling details. This is accomplished by providing stop bars, the top and inside surfaces of which cooperate respectively with horizontally and vertically disposed surfaces of the beryllium copper details that are

fastened to the bottom of the bars. Both the top and the inside surfaces referred to are faced with nylon. The combination stop and guide details are located so that the top surface of the solenoid operated bars lies about $\frac{1}{64}$ " below the pole faces of the pull-down magnet when the bars are fully depressed. This under flush condition is necessary to assure that the pole faces act as the down-stop for the selected card as is intended.

CENTER GUIDES

To maintain the intended vertical disposition of the solenoid operated bars, additional comb-like guides are provided that engage the bars at their center near the bottom edge. Actually they engage the graphitized phenolic linin details that are inserted at the center of the bars. This arrangement of the comb guides provides triangularly disposed guidance that affords both transverts and vertical stability with minimum working friction.

OFF-NORMAL CONTACTS

It is important that all of the code select bars necessary for selecting the required routing card be checked down before proceeding further with a translation and it is equally important that after translation they all be checked up before proceeding further towards restoration of the translator to normal. Off-normal contact springs that are operated by the bars are provided for this purpose. They are mounted at the light source end of the bars as indicated in Fig. 18. They also are illustrated by Fig. 30, and from this illustration it will be noted that off-normal contacts also are provided for the card support bars. These latter contacts close as the bars are operated to start timing of the "no card" timing circuit. When the card support bars are released, their off-normal contacts control the circuit of the card-support lift-bar magnets. This is a desirable arrangement because these magnets are required to supplement the solenoid retractile springs only as the card-support life-bars approach their upper stop position where they engage any cards that may be tilted. Late energization of these magnets prevents the development of high tab impact forces.

LATCHES

The latches are magnetically operated to free the solenoid operated bars and are spring returned to their bar engaging position. Four are provided: two at each end of the bar array, one for the left-hand grouping and the other for the right hand. Accordingly, each has to be capable

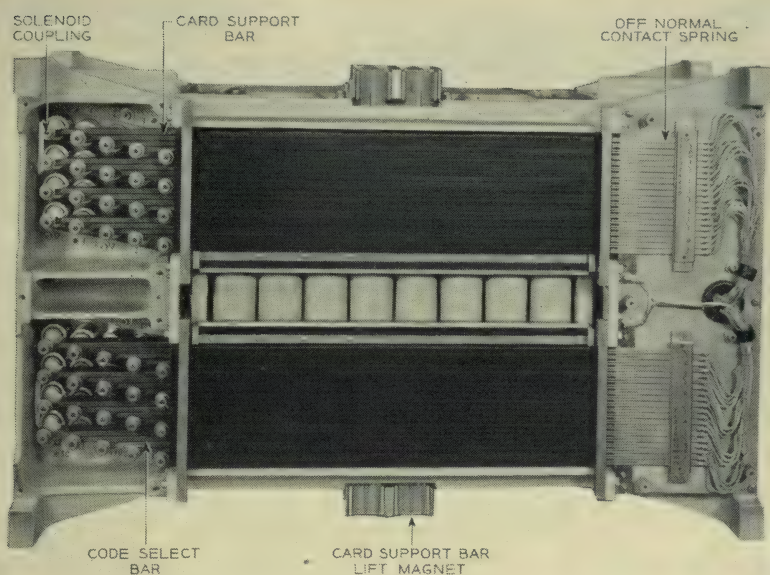


Fig. 30 — Top of dynamic unit illustrating off-normal contacts and solenoid arrangement.

of supporting one-quarter of the weight of a full complement of cards or 25 lb each. Yet the latches have to be fast. This is accomplished largely by providing for "on center" loading as may be discerned by reference to Fig. 31 keeping in mind that the beryllium copper appendages on the bottom of the code select bars engage the support blades of the latches virtually directly above the pivot center. Furthermore, a major portion of the moving member is made from magnesium. A spring pileup such as may be seen in Fig. 32, is associated with and is operated by each latch and the translator circuit is so arranged that a translation cannot progress beyond the latch operate point until all four latches have operated. The same is true when the latches are to be restored to normal, that is, the cycle cannot progress further until all four latches are restored.

The operate time of the latches averages 40 milliseconds, the restoral average being 15 milliseconds.

SOLENOIDS

The main problem here was to provide high-speed operation in the presence of a comparatively large non-operated gap and considerable

load and to do so with a small diameter solenoid. Magnetic cross-fire possibilities born of the close mounting conditions required also had to be contended with. Regarding diameter limitations, it will be recalled that the tabs of the cards are on $\frac{3}{16}$ " centers and that five solenoids are grouped in stagger arrangement. Thus six positions or five spacings have to be considered and, therefore, the diameter of the solenoids has to be less than $1\frac{5}{16}$ ". This is very small considering what the solenoids have to do, but satisfactory performance was achieved by the use of solenoids of which Fig. 33 is representative. The coil tube is made from beryllium copper because of its high resistivity, resistance to corrosion and its ability to resist wear. The main body of the operating plunger is surfaced with a heavy plating of chromium. The plunger then is ground to size, lapped and buffed to provide a very smooth surface. The blade-like extension that is secured to the main body of the plunger for coupling purposes is made from steel. It is hardened, ground and polished to assure long life. The fixed plunger is adjustable axially so that a small but definite operated gap may be provided as otherwise mush-

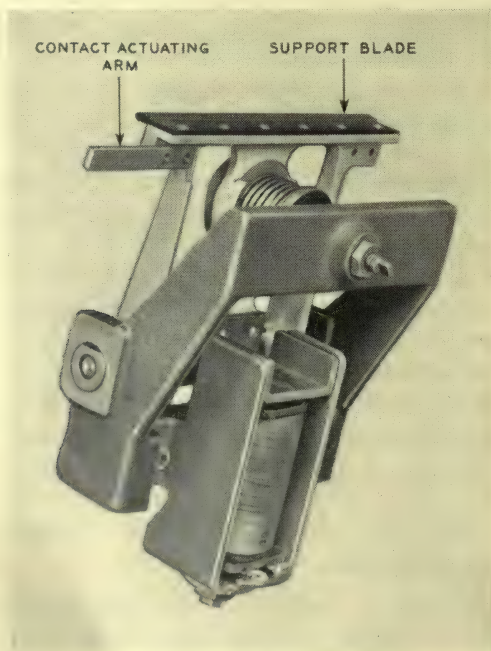


Fig. 31 — Code select bar latch.

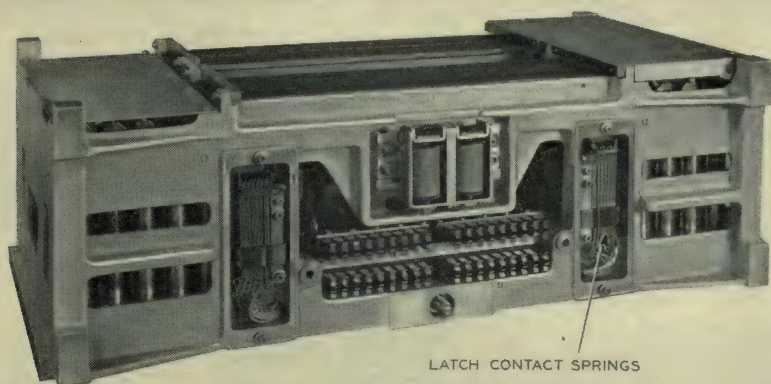


Fig. 32 — Jack side of dynamic unit showing latch contact springs.

rooming due to impact would be experienced. It is possible to maintain a small operated gap because the down-stop position of the solenoid operated bars is precisely controlled by common stop details that engage the solenoid operated bars and thus limit the motion of the solenoid plungers to which they are coupled. The solenoid is provided as a package including the necessary retractile spring and coupling details. One of the coupling details is a rubber buffer. This buffer cushions the impact developed when the bar that is coupled to the solenoid engages the up-stop. At the top of the bar coupling detail another rubber buffer is mounted. This second buffer cushions the impact when the downstop is engaged. A crowned washer is cemented to the first buffer and the assembly provides a slight vertical freedom of the plunger extension in the bar coupling detail. Combined, these features permit sufficient freedom of action to prevent binding in the event one solenoid of a bar acts faster than the other.

The average operate and restoral times of the solenoid are 32 and 28 milliseconds, respectively. The unoperate gap pull averages 325 grams, and the unoperated force of the retractile spring is 200 grams. The solenoids are operated dry to prevent gumming.

The beryllium copper coupling details secured to the ends of the solenoid operated bars provide a slot through which the flattened extension of the solenoid plunger passes. The position control of the solenoids and of the bars is such that side friction at the coupling is virtually non-existent. There is adequate end or longitudinal clearance to permit of unrestrained tilting of the bars such as may be experienced if one solenoid operates faster than the other.

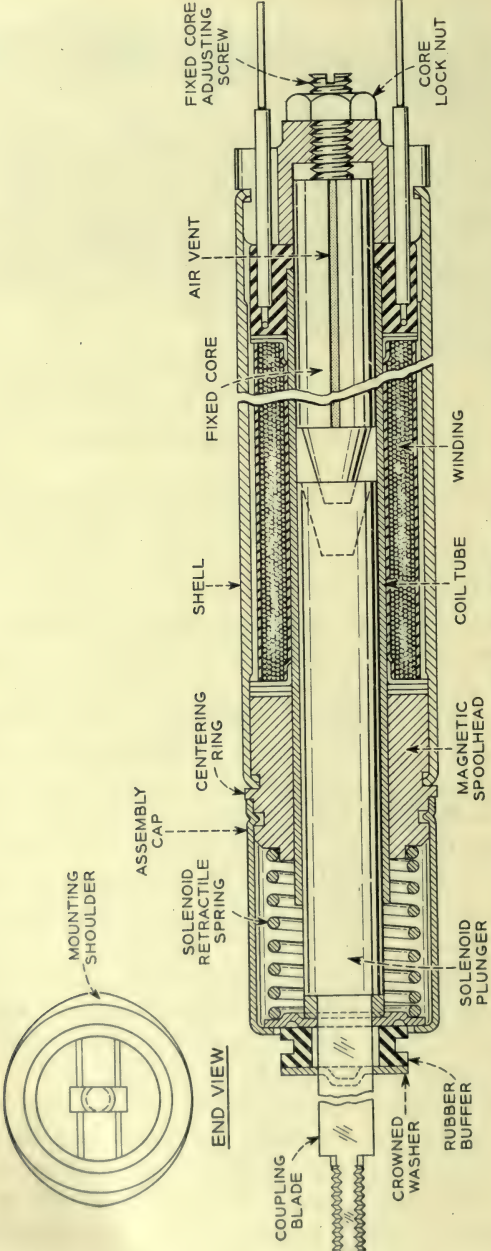


Fig. 33 — Solenoid.

CARD SUPPORT BAR LIFT MAGNETS

It is possible that a translator may be equipped with a group of cards from which all tabs except a very small number common to the group have been removed from one side. To select any card of the group all of the code select bars associated with the remainder of the tabs on the side involved have to be operated. Under these conditions the unselected cards of the group would be unsupported on one side and would, therefore, tilt to the extent permitted by the card guide bar working clearances that prevail. The closest practicable clearances have been established and yet, insofar as the card guide bars are concerned, the lower corner of the cards would be free to drop approximately 0.015" below the nominal bottom surface of the card stack. The combined weight of the tilted cards could be more than the associated code select bars can support; therefore, these bars would be depressed slightly. Actually they could be depressed sufficiently to interfere with the intended free action of the latches when they are called upon to operate and release for restoral of the translator to normal. This could happen, despite the fact that the powerful pull-up magnet would be energized at this time, because the flux shunting action of the adjacent and higher cards reduces the pull on the tilted cards to a very small value. The high cards in rising decrease the chances of the tilted cards following. Therefore, supplementary magnets have been provided that assist the retractile springs of the card support bars in lifting the tilted cards. These magnets are powerful enough (6.6 lbs.) to straighten up the tilted cards.

The form of these magnets, which have been called card support bar lift magnets, is illustrated in Fig. 34. It will be noted that, like the latches, they are constructed as units to permit convenient servicing in the field. Two coils are used for each magnet because in this way standard comparatively small diameter coils can be used and thus the magnets may be mounted more readily in the space available.

214A SELECTOR

All of the dynamic components, that is, the code select and card support bars, the solenoids, the latches, the card support bar lift magnets and the off-normal contacts are combined in a subassembly that is arranged on a plug-in basis. This combination, which as a convenience includes the pull-down magnet and of necessity many supporting details, is illustrated by Figs. 30 and 32. It is manufactured as the 214A selector but commonly is called the dynamic unit. It may readily be removed from a translator and placed on a portable elevator table such

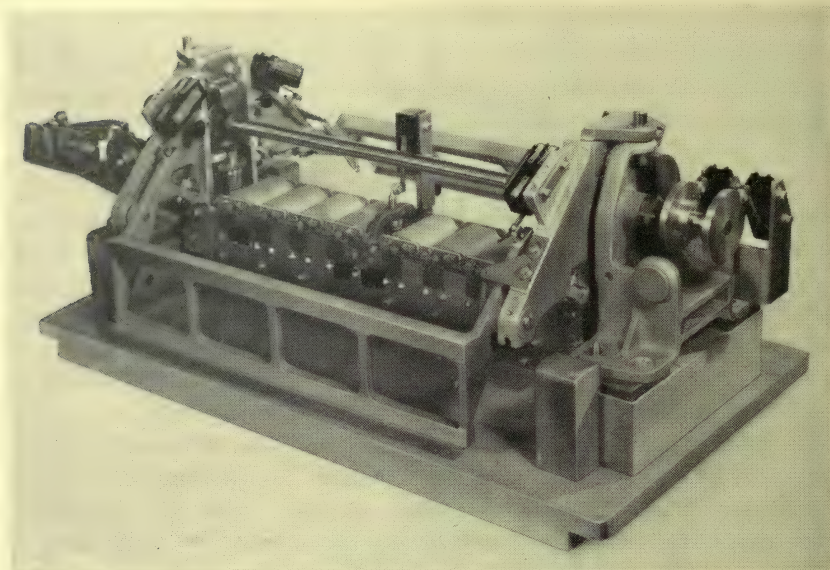


Fig. 34 — Pull-up magnet and lift superstructure.

as is depicted by Fig. 35. Accordingly, when the dynamic unit needs to be serviced, it may be taken to a location suitable for the purpose and meanwhile the translator may be continued in service by substituting a spare dynamic unit.

REMOVAL OF THE UNIT

To remove the dynamic unit, the cards first either have to be removed from a translator or in some way they have to be supported clear of the code select and card support bars. In some cases it will be possible to service the unit in a few minutes, and in such cases, should the cards have to be removed, their handling would constitute the major effort. The pull-up magnets could be energized in which event the cards would be suspended clear of the solenoid operated bars thus permitting the removal of the unit. However, conditions might be such that the pull-up magnet thus would have to be kept energized for a comparatively long time and this would be undesirable because the current drain is heavy. In any event, suspension of the cards by the pull-up magnet for this purpose would be hazardous because should the circuit be interrupted only momentarily, the cards would fall and considerable damage no doubt would result. Accordingly, mechanical means were provided whereby the cards may be elevated quickly and held suspended safely

for an indefinite period. This was accomplished by providing hooked end bails pivotally mounted on the pull-up magnet structure and so arranged that they may readily be swung into a position where their hooked ends enter the top notches in the end edges of the cards. A manually operated crank is provided whereby the bails and hence the cards may be elevated. The mechanism employed is shown in Figs. 27, 28 and 34. It will be noted that the bails are arranged to act much like ice tongs, that is, the heavier the load the more securely they grip.

CLEARING A "CARD CRASH"

It is possible, under certain conditions, involving double trouble, for the pull-up magnet circuit to be broken while the latches are operated clear of their code select bar supporting position. If this happens, the card stack will fall and unless the translator is very lightly loaded, the weight of the card stack will overcome the solenoid retractile springs and all of the cards will move down to the pull-down magnet pole face level. Such a condition has been dubbed a "card crash". If it occurs, the pull-up

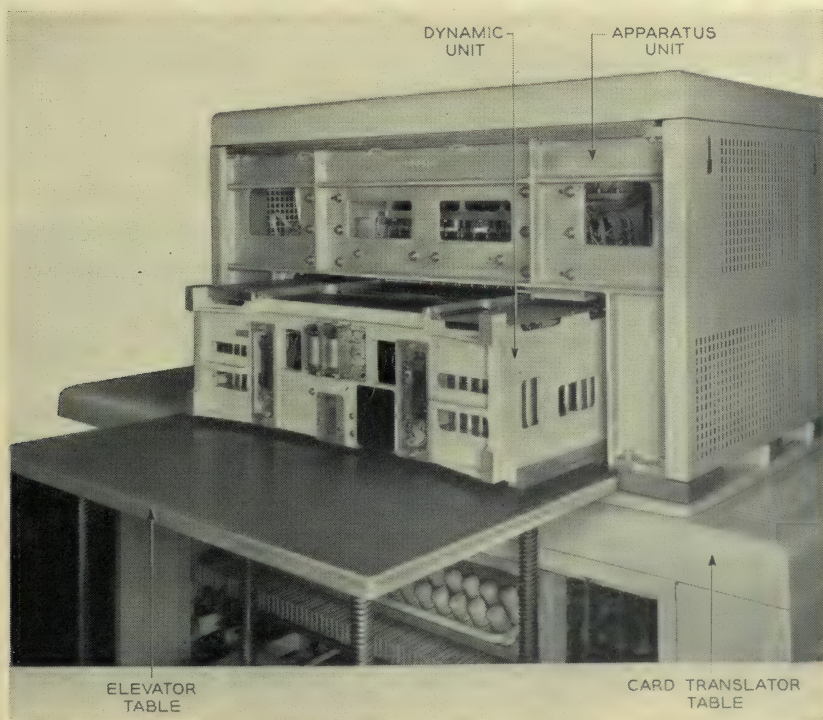


Fig. 35 — Portable elevator table.

magnet will be incapable of lifting the card stack because of the materially increased airgap and, therefore, the translator will be rendered inoperable. However, the bails provide a convenient means for clearing the trouble because the top notches in the cards are so located and are of such a size that they may be engaged by the bails even when the cards are at the pull-down magnet level and the lift motion is sufficient to permit the latches again to assume their card stack supporting role.

REMOVAL OF CARDS

To remove the cards, the pull-up magnet must be lifted sufficiently to permit of the cards being raised above the lower card guide bar. In addition, the upper card guide bar must be lifted clear of the stack. The crank referred to may be used to effect the required action. The upper card guide bar is raised by rotating the crank beyond the position required to raise the pull-up magnet whereupon a cam of the internal peripheral shell type engages mechanism adapted to lift the guide bar while the pull-up magnet dwells in the elevated position already effected. The mechanism referred to is illustrated schematically in the Figs. 27 and 28.

Cams, locking pawls, and associated circuitry cooperate to prevent operation of the crank while the translator is in service, to limit to the necessary extent the amount that the crank may be turned when removing the dynamic unit or alleviating a card crash and to permit of its being turned sufficiently further to elevate the upper card guide bar to clear the cards when they are to be removed or are being loaded. Study of related figures will make clear the mode of operation of the elevating mechanism.

PULL-UP AND PULL-DOWN MAGNETS

These magnets (see Fig. 18) are of the eight coil common pole type. They differ principally in that the cores for the coils of the pull-up magnet are laminated to provide for fast operation as is required. The operate and release timing of the pull-down magnet is such that its cores need not be laminated.

It is of vital importance that all of the coils of the magnets are effective as otherwise some of the cards may not be elevated when the latches are called upon to operate, or the cards may not fall reliably. Accordingly, a slave relay is used in series with each of the coils and if a relay fails to operate, thereby indicating possible magnet coil failure, the trouble recorder is called into action and steps are taken to clear the trouble.

The pull-up magnet is capable of exerting a pull of 350 lb when the airgap is 0.020" which is approximately 25 per cent more than the normal operate gap and since the cards weigh approximately 100 lb. it will be seen that a reliable operate margin has been provided. The performance capabilities of the pull-down magnet were developed in the discussion of its effect on the drop time of the cards. The breakaway pull of this magnet is approximately 1,000 lb.

LIGHT SOURCE

The light source comprises the projection lamp, the light beam modulating wheels, the wheel driving motor, and the two first surface mirrors, as illustrated by Fig. 18. The mirrors direct light from either side of the major plane of the lamp filament via the dual collimating lenses and the card stack formed light tunnels to the photo-transistors. Various adjustment and supporting details are combined in a subassembly illustrated by Fig. 36. The subassembly is shown mounted on a fixture that is used during manufacture.

The lamps usually can be interchanged without readjustment of the mount or the associated components. At the most, all that has to be done is to orient the lamp so as to obtain reasonable alignment of the major

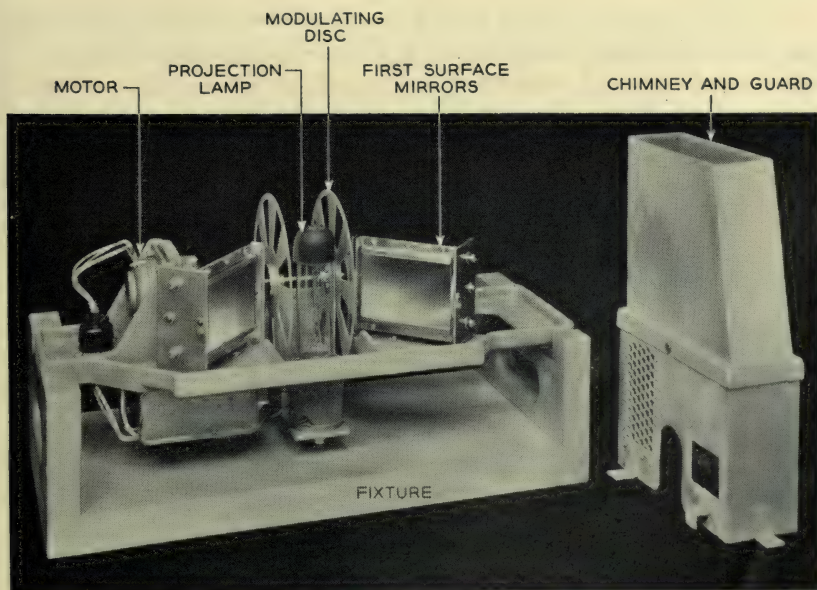


Fig. 36 — Exciter lamp assembly.

plane of its filament with the longitudinal center line of the translator. The combination chimney and guard that is provided as part of the subassembly includes a peep-hole to facilitate filament orientation.

The lamp, although operated at reduced voltage, generates considerable heat. Because of this, its guard and chimney are treated with a cotton flock finish which prevents direct contact with the metallic base and thus the guard may be touched without discomfort.

The first surface mirrors are mounted so that they are accessible for cleaning.

COLLIMATING LENSES

The light level at the photo-transistors may be lessened because of the accumulation of foreign material on the surfaces of the lenses. Because of this, the collimating lenses have been mounted in a frame (see Fig. 37) which makes it convenient to remove them simultaneously for cleaning and to replace them without readjustment. The photo-transistors are red sensitive and, therefore, the focal length had to be determined on this basis.

CARD CAGE

The cards are mounted in a cage, as illustrated by Fig. 25. The partitions are made from magnetic iron. They tend to repel the cards that are mounted adjacent to them when the pull-up and pull-down magnets are energized. This is beneficial when a card close to the partitions is selected, as it gives better assurance of the card dropping reliably. However, as has been mentioned, it is desirable to mount an uncoded card adjacent to the partitions because the first and last card in a com-

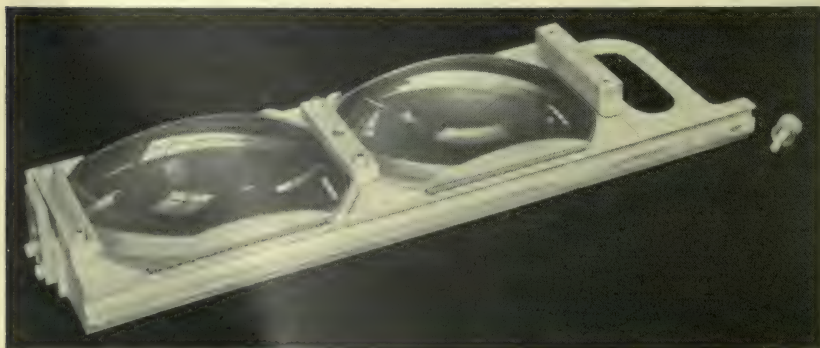


Fig. 37 — Collimating lenses.

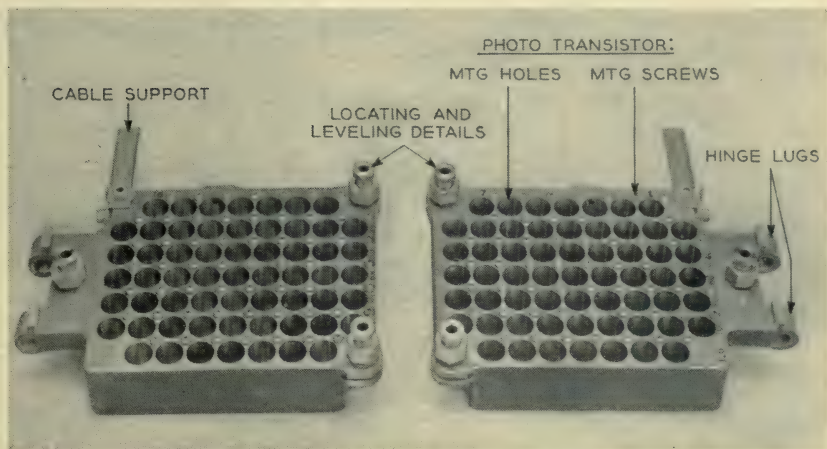
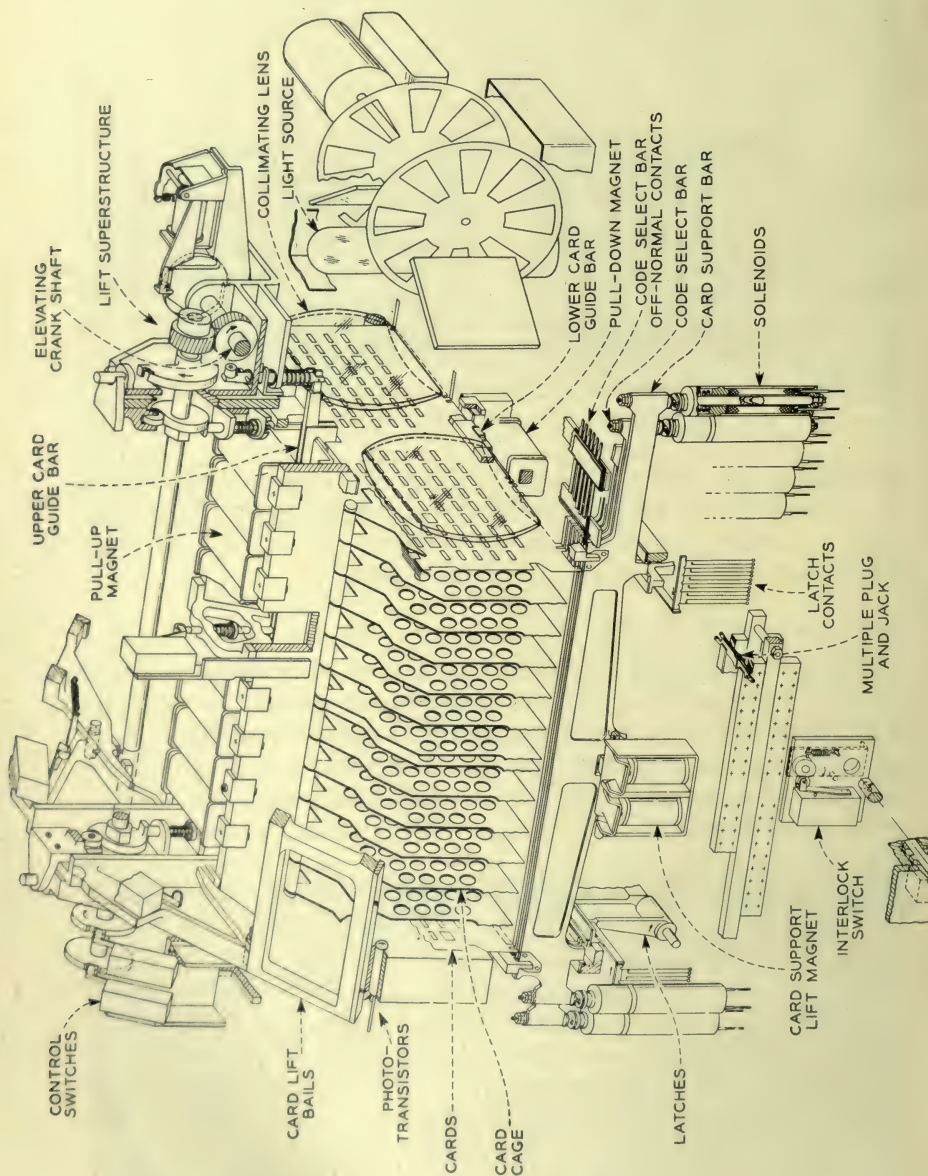


Fig. 38 — Phototransistor mounts.

partment are not as free acting as the others. The holes in the partitions, it will be noted, are round. Round holes were used merely as a convenience in manufacture. They are of sufficient size to clear the light beams that pass through the card formed tunnels.

PHOTO-TRANSISTOR MOUNT

The photo-transistors, consistent with the left and right-hand grouping of the holes in the cards, are mounted in left and right-hand groups of fifty-nine each. The mounts used comprise frames that are precisely machined, necessary because the activating light beams have to be concentrated on a very small piece of germanium and many variables are involved. These frames normally are secured to the translator framework by means of three special mounting screw subassemblies that provide means for positioning and leveling to obtain optimum light distribution over the whole field of fifty-nine photo-transistors. However, it was recognized that the focusing lenses of the photo-transistors should be readily accessible for cleaning and that it should be possible readily to view the light beams transmitted by the card formed tunnels as for instance by placing a piece of ground glass or translucent paper near the end of the card cage. In this way preliminary adjustment of the first surface mirrors, three dimensional adjustment of the light source, etc., can be effected quickly. Accordingly, the special subassemblies include screws that may readily be removed without affecting the position or leveling adjustments. As a convenience, loose hinges also are provided so that after the mounting screws have been removed, the



mounts may be swung aside as gates to gain access to the photo-transistor lenses for cleaning. The photo-transistor mounts are illustrated by Fig. 38.

COVERS

Covers are provided for the four sides of the translator and its top. These covers may be dismantled quickly when servicing is necessary. The base casting is cut away beneath the covers so as to form air passage ways to its interior. The covers are perforated near their top and in this way a chimney effect is created which is beneficial to cooling.

4A AND B APPARATUS UNITS

All of the static elements, that is, the light source, the collimating lenses, the card cage, the pull-up magnet superstructure, the photo-transistor mounts, the terminal block, and the multi-terminal jack sub-assemblies, the framework, the covers and sundry details are combined in a subassembly known as the 4 type apparatus unit.

COMPREHENSIVE SCHEMATIC

The over-all appearance of the working translator is disclosed in Figs. 12, 13, 39 and 40. The principal component groupings entering the assembly have been discussed at some length and have been illus-

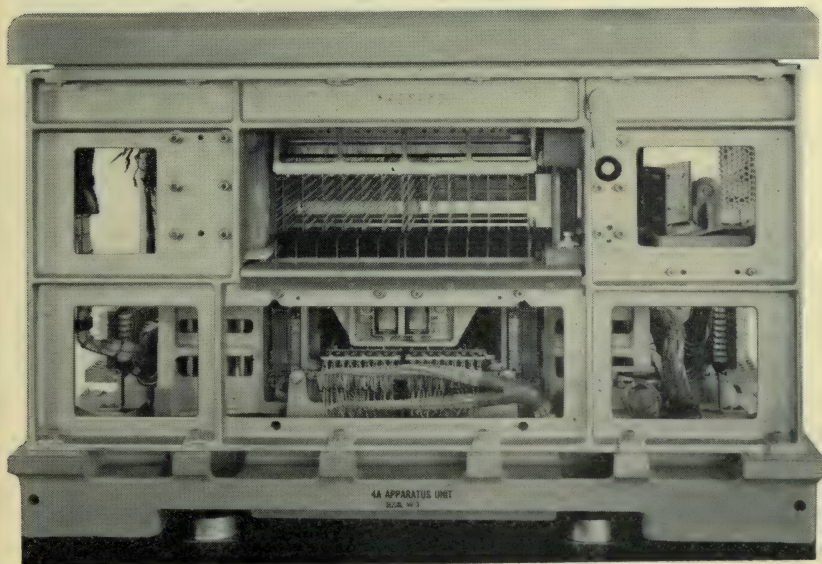


Fig. 40 — Translator (card access side).

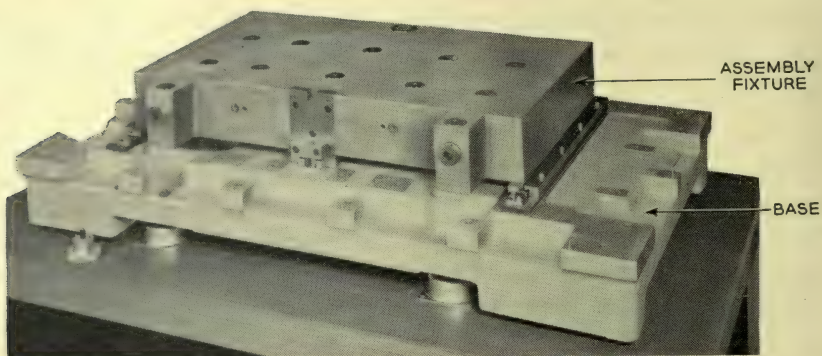


Fig. 41 — Base and assembly fixture.

trated in other figures. However, the combination of these principal components needs to be illustrated more specifically. Fig. 39 serves this purpose.

WORKING TRANSLATOR

Fig. 40 is a view of the working translator showing how the supporting structures knit together the principal components illustrated

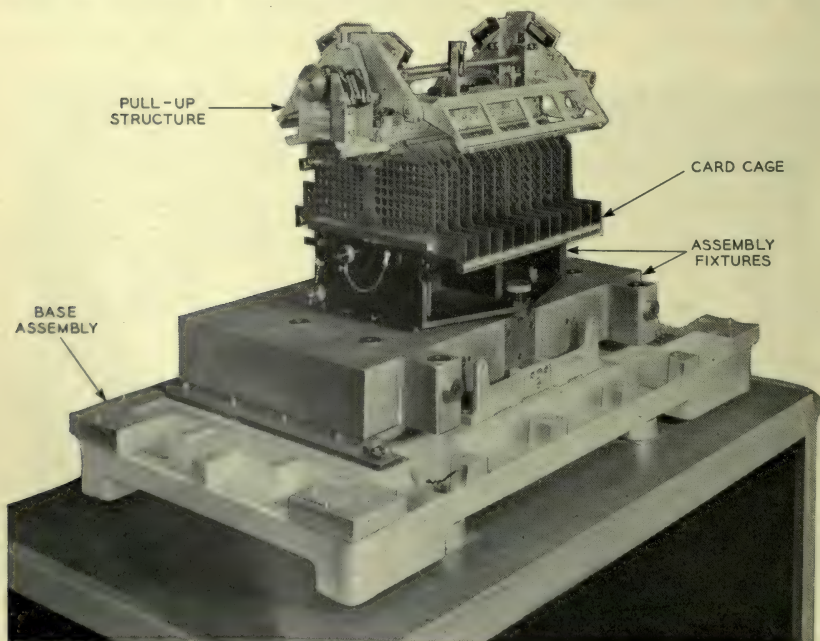


Fig. 42 — Base and assembly fixture together with card cage and pull-up superstructure.

in perspective by Fig. 39. It supplements Figs. 12 and 13, which depict the translator with its covers in place. The coded translator (1A) is 38" long by 21- $\frac{1}{2}$ " wide by 27" high. Its weight is approximately 410 lb. The weight of the dynamic unit is approximately 110 lb.

BASIC PLAN FOR CONSTRUCTION

As the development progressed, it became apparent that unusual and exacting manufacturing problems would be encountered which could be solved to best advantage by providing design features that would facilitate alignment of the light tunnels, the optical elements, the card select and card support bars with respect to the tabs of the cards, etc. Accordingly, the design was worked out so that the assembly could be built up around a base casting provided with reference surfaces and locating pins adapted to be used with a surface plate type of fixture that could be mounted in the space normally occupied by the dynamic unit and upon which supplementary fixtures could be mounted for

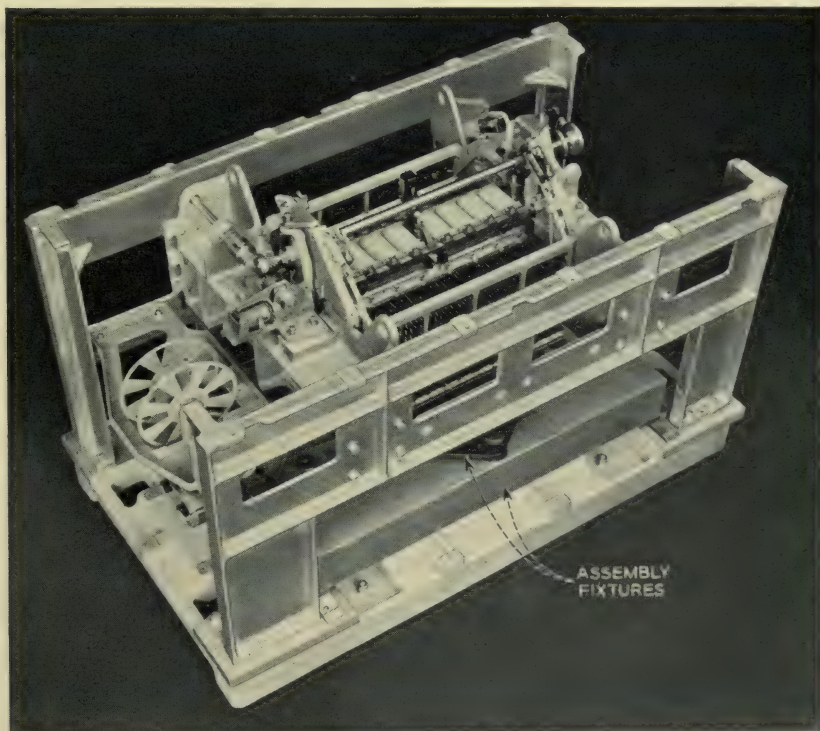


Fig. 43 — Base and plug assembly fixture, card cage and frames.

aligning the various components. With fixtures of this kind virtually all assembly can be made without resorting to measurements. The design of the translator and the fixtures were, therefore, carried forward concurrently.

An initial step in fabricating the card translator with the aid of these fixtures is illustrated in Fig. 41, in which the basic surface plate is shown mounted on the base casting. Fig. 42 shows another step in the fabrication of a translator. It will be noted that in this instance a supplementary fixture had been added which supports the card cage and the pull-up magnet superstructure. Later the side frames are added and these too are located by means of the fixtures as is indicated by Fig. 43. It will be understood that many additional fixtures are used in completing the job and that one of the most important of these is a fixture which locates the gates in which the phototransistors are mounted. However, by the time this fixture is used, the assembly has built up to a point where the fixture used is obscured by the side frames and other details. This method of construction has proved to be very effective as is evidenced by the fact that after assembly there have been no cases where the performance requirements could not be met readily because of misalignment of components.

The development of nationwide dialing is the result of close cooperation of many people in the American Telephone and Telegraph Company, the associated Telephone Companies and particularly for the card translator, the Western Electric Company, as well as of many people in Bell Telephone Laboratories. All who have had a part in this development, which may justifiably be considered a milestone in the development of the switching art, can view with pride their accomplishments. This development, serving as the means for connecting on a nationwide basis all the customers' lines into one vast automatic switching network, is truly a major contribution to the telephone art.

Development and Manufacture of Electroformed Conductor for Telephone Drop Wire

By A. N. GRAY and G. E. MURRAY

(Manuscript received June 15, 1953)

Telephone drop wire is that familiar black overhead wire which brings the telephone service to the home. It is a parallel pair of conductors separated and positioned in an extruded insulation, covered by a cotton serving and jacketed with a neoprene compound of tire tread-like qualities. In the past, a cast copper jacketed steel ingot, rolled and drawn to size, has been used to provide a wire that combines high strength and good conductivity. In order to assure more than a single source of supply and to provide improved mechanical and electrical characteristics, a completely new plant has been constructed for continuous plating of steel wire at completed size. This process provides a stronger, yet smaller and, therefore, less costly wire than was possible previously. Plating is done at 100 feet per minute on 25 wires simultaneously. The conductor is then processed as formerly to provide the neoprene jacketed drop wire.

HISTORICAL BACKGROUND

It has been customary in manufacture to apply a lead and a brass plate to drop wire conductor to secure good adhesion to the insulating compound. The brass provides the adhesion while the lead prevents attack on the copper by the sulfur in the rubber. In 1941, two tandem lead and brass plating machines were placed in service at the Point Breeze Works of the Western Electric Company to apply these coatings on a production basis. Their successful operation proved that electrolytic deposition of two metallic coatings in tandem at high speed was commercially practicable. It was not difficult to imagine the addition of a copper plating section to deposit copper, in addition to lead and brass, on a steel wire as a combined operation. The technical problems involved,

however, were of considerable magnitude, and their successful solution required a substantial amount of investigation and development.

Such a process appeared to offer a number of attractive features. The maximum strength of the steel core could be utilized since no compromise in physical properties would be required to permit rolling and drawing. A highly uniform cross section could be secured which would provide continuous copper protection for the steel core against corrosion. Since the steel core wire would be a standard commercial item available from a number of manufacturers, alternate sources would be available to assure continuity of supply.

There were also substantial economic inducements. The higher strength wire of uniform construction would reduce the cost of trouble calls to the Bell System. It appeared that a plant could be designed which would require no more labor to operate than was required for the existing lead-brass plating operation. By starting with a steel wire of uniform and circular cross section and applying a uniform copper jacket, the desired physical and electrical requirements could be met by a conductor $2\frac{1}{2}$ thousandths smaller in diameter, which would, in turn, reduce the overall dimension of the finished product. Although this reduction might appear small, the very large footage of wire required indicated a saving of a half-million pounds of copper a year, and a combined saving of steel, copper, rubber, cotton and neoprene, all strategic materials, of a million pounds a year. Combined savings to the Bell System Operating Companies and Western Electric Company were estimated at better than one million dollars a year.

World War II prevented further work until 1946 when the project was reopened and methods of obtaining heavy copper deposits investigated on a laboratory scale. Initial developments showed promise and late in the year, a separate development laboratory was set up to investigate various electro-chemical problems and to develop information on which a pilot plant could be designed. The term "Electroforming" was first applied at this time because it was apparent that the process was not to be one of electroplating in the ordinary sense but rather was to substantially change the physical and electrical properties of the wire. In other words, the terminology was intended to differentiate between utility and what are usually decorative or protective functions.

A pilot plant was built and installed which operated successfully the first day it was placed in operation. The results obtained on the pilot machine exceeded expectations. The limiting current density for the acid copper plating solution determined in the laboratory was 1,000 amps./sq. ft., whereas 2,000 amps./sq. ft. was realized on this machine. This in-

crease in current density meant a rate of deposition double that which had been predicted.

On this pilot machine the operating limits of the various plating and cleaning solutions were established. Methods of control, materials of construction and design features were evaluated. A field trial lot of 200,000 linear feet of drop wire and numerous samples of wire were processed for examination and design approval.

Some time was spent in making the basic plant decisions, evaluating pilot plant experience and carrying through the numerous special investigations required to guide the engineering design. The detailed engineering and drafting were begun and firm orders placed for equipment.

DESIGN PREMISES

Certain design premises became apparent from experience with the pilot plant. The nature of the process dictated continuous operation on a three-shift, seven-day basis. To secure such operation, it was necessary to duplicate certain critical facilities, use the largest practical reel size for maximum wire run time and to employ great care in the design of the wire handling equipment to minimize wire breaks. The second premise was low maintenance. This required that the materials in contact with the various chemicals could be selected only on the basis of extensive corrosion tests. This involved the study of many of the grades of stainless steel as well as the rarer metals and the broad field of plastics and elastomers to select suitable materials for tank lining, machine parts, and piping.

Substantially automatic operation was set up as another design objective. This, of course, involved the isolation of the factors requiring control and selection of the most suitable means. The safeguarding from waste of valuable solutions was a fourth consideration. Spare tank capacity was provided in case any of the storage tanks developed leaks and had to be repaired. Facilities were required to recover electrolyte carried out from plating operations by the wires themselves. Means had to be provided for recharging and reconditioning the various electrolytes. Still another premise was the permanency of solution: no dumping and replacing of plating solution was contemplated. And, of course, safety to personnel was a must. In addition to the usual hazards from acids and alkalis in a chemical plant, accidental mixture of electrolytes could give rise to highly poisonous gases. This necessitated the selection of highly reliable piping materials and the provision of adequate employee protective devices and routines.

BASIC PROCESS

The starting point in the process is commercial improved plow steel wire 0.0336" in diameter delivered on 450-pound reels. Twenty-five strands pass through the plating machine in parallel at 100 feet per minute. After suitable cleaning, approximately 2-1/2 thousandths of copper plus a thin plate of lead and brass are electrolytically deposited on the wire. To apply this deposit requires nine different electrolytes, approximately 80,000 amperes at 5 volts, and a 600-ft machine with a wire span from supply to take-up of 850 feet.

The plating portion of the machine is a relatively simple structure consisting of a long trough containing plating cells alternating with contact rolls. The electrolytes are pumped into the plating cells through which the wire passes, cascade into return troughs, flow back to reservoirs and are continuously recirculated. The contact rolls position and propel the wires through the machine and serve as the means of making electrical contact.

The general structure of the machine is uniform throughout, only the material changing to fit the chemical requirements of the electrolyte in the particular section. The wire, in passing from one electrolyte to another, travels through washing and wiping facilities mounted in the troughs to prevent contamination of electrolytes and reduce dragout of valuable solutions. The finished wire, controlled to specified conductivity, is then taken up on reels ready for insulating.

THE BUILDING

The building is 91 feet wide by 340 feet long, of brick and steel construction, in keeping with Point Breeze architecture. It was specifically designed to fit the process. The first floor is given over to wire supply and take-up, electrolyte mixing, pumping and conditioning and material storage. All plating operations take place on a mezzanine. The upper portion of the building is divided into three bays by a pair of lengthwise partitions. The center section contains the two plating machines, each machine being constructed in the form of the letter "C", the two "C" shaped machines being placed face to face. The outer bays contain the rectifiers, electrical controls and heating and ventilating equipment. The floor of these bays is steel grating. The partitions prevent the entrance of any vapors released by the heated plating solutions, and the ventilating equipment forces a steady stream of clean and tempered air downward over the electrical facilities and through the grating into the first floor area below.

The ventilating equipment draws in fresh air from the roof, heats it

if required, and distributes it down the electrical bays by means of overhead ducts. The system operates under modulating temperature control with 100 per cent fresh air make-up. After reaching the first floor, the air then flows upward through the center section of the building, past the plating machines and is exhausted through the roof by fans at the rate of 160,000 cubic feet per minute.

Power for the building is supplied from a substation located in the south electrical bay. A pair of underground cables bring in energy at 13,200 volts to high-voltage switches for distribution to two identical, 1500 kva, three-phase transformers where it is stepped down to 480 volts before entering low voltage switch gear for distribution about the building. The electrical system is designed so that the entire plant load can be supplied by either of the high-voltage cables and for a short period of time by either of the step-down transformers. Steam, city water and house water are furnished through an underground tunnel.

In addition to electric power, the building is furnished with 150 lb./sq. in. steam for process and heating, 90 lb./sq. in. compressed air for control instruments and various process functions, city water and house water. A sanitary sewer for the washrooms plus separate acid and alkali sewers of chemical resistant pipe have been provided from the pits where the electrolyte storage tanks are located. These latter sewers are for emergency use only. All solutions are reconditioned and recirculated in normal operation and as the plating solutions are permanent no disposal plant has been deemed necessary.

Lighting throughout the plant is furnished by incandescent lamps in simple porcelain reflectors, designed to give a minimum of 17 foot candles at all locations. This is supplemented by fluorescent fixtures at the take-up stands because experience has shown that appearance of the wire on the take-up reels is an important indication of the quality of the plate. The entire interior of the plant, as well as all machinery, is finished with vinyl base paint. Extensive tests were made of various types of chemical resistant finishes and the vinyls were found to be the outstanding performers.

For moderate service conditions, metal surfaces were wire brushed and washed down with cleaning solution before the wash type primer was applied. For severe service, the surfaces were cleaned by sand blasting before priming. The customary primer coat was then applied and several top coats, the number depending upon the severity of the service.

WIRE HANDLING EQUIPMENT

Tonnage-wise the steel core wire is the largest, if not also the most important, single item of raw material procurement. In searching the

market for a high-strength, bright finish steel wire which would be readily available from a number of supply sources, it was found that improved plow steel wire as regularly used in wire rope manufacture would meet the proposed requirements and could be purchased under the several wire mills' own control specifications and tolerances, giving Western the reliability of an established commercial supply without the price premiums for a specialty wire.

In only one important particular was it found expedient to deviate from the steel mills' standards as regards the steel core wire, and this specifically concerned the packaging of the wire to facilitate subsequent handling in our plant. It has been customary for the steel mills to ship this type of wire to customers in paper-wrapped bundles which are block-wound, catch-weight coils, shaken down and bound with soft iron tie wire. Because of the coil-forming action on the draw-blocks, the steel wire mills have found that bundles containing more than about 250 pounds of 33-mil wire cannot be made without greatly increasing the dangers of tangling and breaking when wire is payed out from the bundle.

An economic study of wire handling in both the electroforming plant and the subsequent insulating and jacketing departments showed that a 450-pound steel wire package free of splices to avoid excessive scrap, cut-over and reel handling losses was most desirable. It was found impractical to put this much core wire in a single bundle with any assurance that the wire could be payed out of the coil without too many breaks from snarling, tangling and fouling of the wire on the pay-off stands.

After extended negotiations with the suppliers, decisions were made to obtain the core wire on reels, in order to insure reliable pay off from supply units of the weights required. Agreements were reached to handle the wire on returnable reels which the mills provided to suit their own winding equipment. No unwinding problem exists at Western Electric because the supply reel is not rotated to unwind.

The nature of the electroforming and electroplating operations renders it impracticable to stop a running wire to replenish an exhausted supply. Consequently, it was necessary to provide for splicing the inner end of the supply wire paying off onto the outer end of a standby supply. Where the supply reels are revolved to unwind and remove the wire, such provision calls for compensator loops or accumulator towers to accumulate temporarily enough feed-out wire to keep the machine going while a splice to the new supply is being made.

Such accumulator devices require a large amount of space which was not economical to provide. They demand that an operator be precisely on the spot to make the splice before the limited feed-out accumulation

is exhausted, an impractical demand considering the small number of operators assigned to the plant. Furthermore, rotating supply reels of the size under consideration require special feed-off drives and controls to maintain a pay-out drag of an order to establish the range of operating tensions demanded in the plating lines. To circumvent the objectionable features of the rotating supply reels, provision has been made to take the wire off "over end" from a stationary reel, lying flat on one of its heads, so that the inner (or tail) end of the wire paying off can be spliced to the outer (or starting) end of a stand-by at any time before the pay-off length runs out. To this end arrangements were made with the steel core wire suppliers to bring the inner end of the wire to the outside within the reel heads, allowing enough free length for making the splice.

Each of the twenty-five wire channels on each of the electroforming machines is provided with a dual supply stand holding two reels in the immediate vicinity of each other, one the pay-off and the other a stand-by (Fig. 1). When placed in position on the supply stand, the upper heads of both reels are set up with flyers to guide and tension the off-coming wire, and facilitate automatic transfer from one reel to the next. The

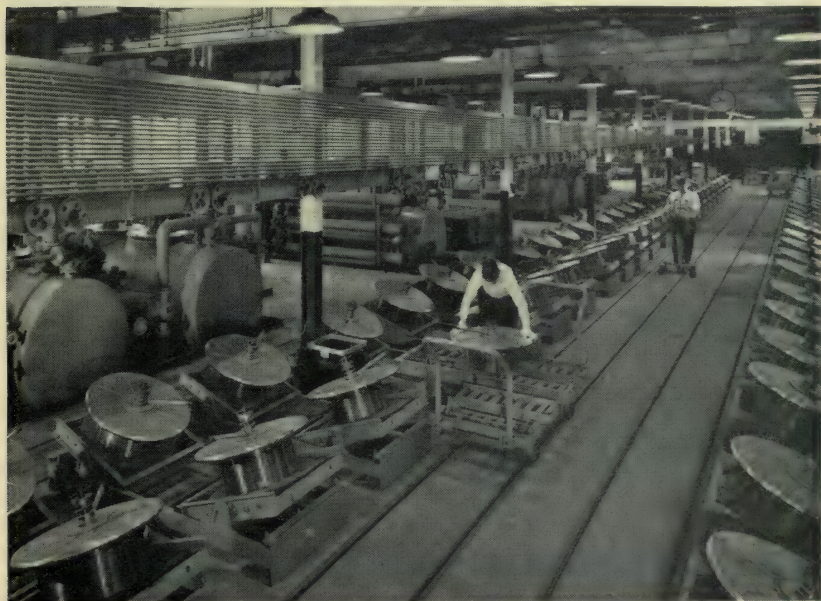


Fig. 1 — Steel wire supply stands, two for each of the twenty-five channels on a machine, permit continuous operation. A 450-lb. spool of wire is being positioned by the nearest operator while behind him another operator is electric welding a wire joint.

flyer rotates about an extension of the reel axis and is lightly restrained with a friction drag-brake which prevents flyer overrun and puts just enough tension in the lead-off wire to keep it from flying wild and tangling. The nature of the over end pay-off causes too wide a change in tension as the supply reel is depleted so the flyer braking force is limited and an auxiliary brake sheave is used to supply most of the back-drag on the wire and create a more stable and uniform approach tension to the machine. From this sheave the wire passes through guide tubes to a fanning section which arranges the wires in proper order and guides them to the supply capstan on the mezzanine at the entrance to the plating line (Fig. 2). The supply capstan is of large diameter and strongly electromagnetic and the wire is snubbed to a 140-degree wrap on it to minimize wire slip and creep. The supply end capstan is the speed governor for the entire machine. It is positively driven with a shunt-wound dc motor regulated to hold capstan speed within $\frac{1}{2}$ of 1 per cent over the adjustable wire speed range of 80 to 120 f.p.m. (Fig. 3).

The twenty-five wires of each machine are carried through the plating line in spaced relationship by passing alternately over and under a series

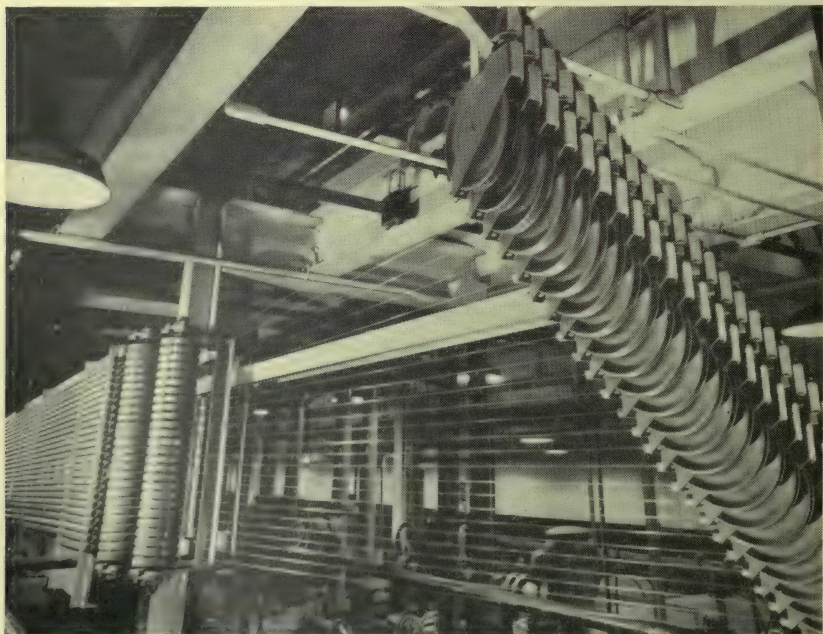


Fig. 2 — String-up of the twenty-five channels of wire carries them from supply spools, left, through guides and sheaves to the mezzanine location of plating operations.

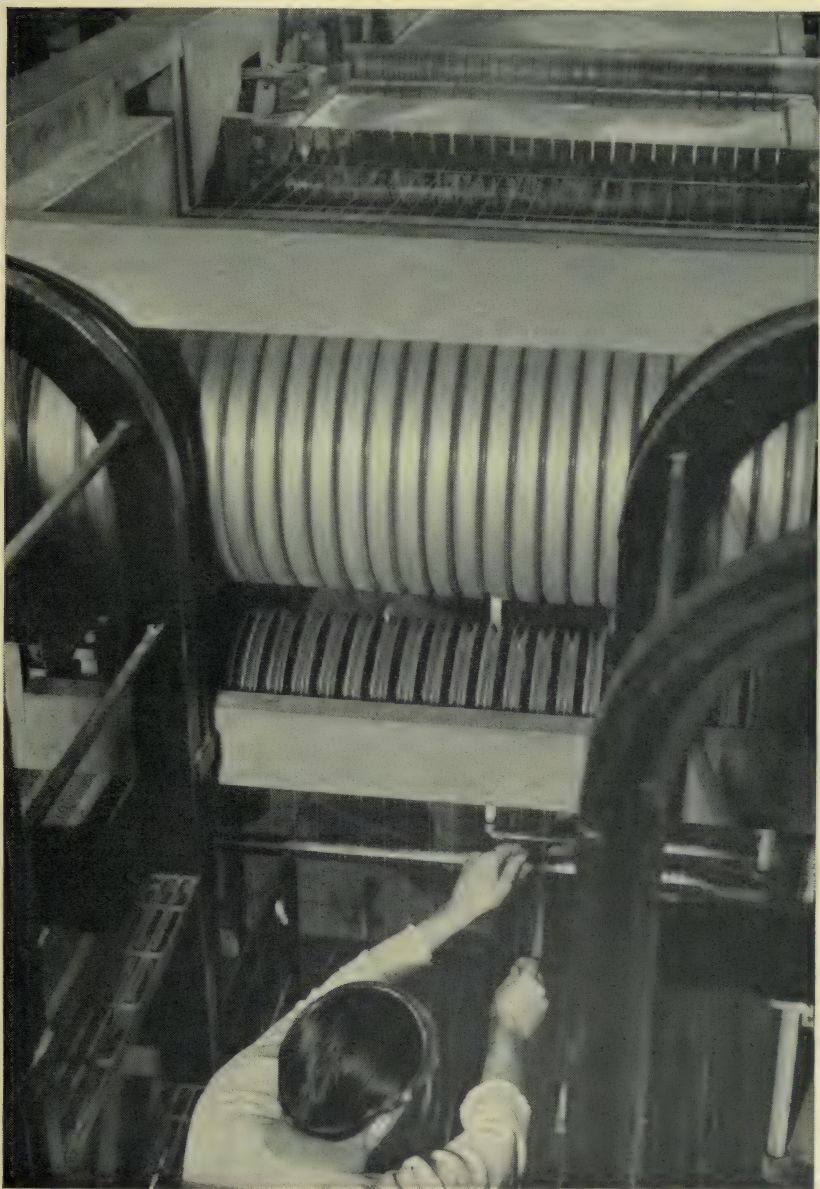


Fig. 3 — Up and over the capstan the steel wires are drawn from the supply position on the first floor to the initial wire preparation baths to begin the 600-ft. journey through the electroforming process.

of grooved rolls. There are a hundred of these rolls in each plating line and all but a very few of them are equipped with contact brushes to supply cathode potential to the wires in the cells. Because of bearing, roll seal, and brush friction, these rolls offer considerable resistance to turning; this frictional resistance is subject to rather wide variations under service conditions. To limit and control the rate of wear on the roll grooves, and to prevent a build-up of excessive wire tensions in the plating lines, all rolls are positively driven at constant speeds. The rolls are grouped on eight separate drive units, each powered by its own motor, so that each carries a definite portion of the total roll load. The motors are geared in and speed regulated by resistances in the motor control circuits to drive all the contact rolls a little faster than wire speed, each successive group of rolls being also driven slightly faster than the preceding one. This roll overdrive, as a result of wire-to-roll friction, produces a closely controllable pull on each individual wire since the only service factor affecting roll pull is the total wire tension (the consequent wire-to-roll pressure) prevailing in any given part of the plating line. The roll pull on the wire increases wire tensions toward the entry end of the plating line so that the point of highest wire tension is at the supply capstan recess, but the total increase is moderate because most of the roll pull is absorbed in overcoming wire drag through the cells.

Wire tensions on the approach side of the magnetic supply and capstan are held low in the interest of safety, efficiency and economy. Those on the recess side of the capstan are higher, which means that the capstan is actually pulled by the wire. So that the motor on the capstan does more than just a braking job, it is made to drive the first group of contact rolls, all those in the preparation leg. In this way the supply capstan motor is loaded about the same as other motors in line so that it operates at a comparable point on its characteristic curve, and performs in substantially the same way as the other motors. It is important that this motor be positively loaded so that it "motors" instead of "generates" because this motor drives a tachometer generator which controls the entire drive.

Most of the contact rolls are in the acid leg of the plating line. They are hardest to drive because of their close spacing and the large number of brushes they carry. To safely limit drive chain tensions and produce the desired roll overdrive gradients the acid leg rolls are grouped in six drive units each with its individual motor. These are matched motors with similar load characteristics and they are geared in at comparable

points on their characteristic curves so that they perform alike under speed and load fluctuations.

The finishing leg rolls and the take up capstan are driven with a single motor identical to the motor driving the supply capstan and the preparation leg. The take-up end capstan adds only slightly to the wire tension over and above that produced at the take-up spools. For that reason the take-up capstan alone does not load the drive motor enough to make it perform like the other motors under load and speed changes. To build up the load on this motor to where it performs at a point on its characteristic curve corresponding to the other motors and "follows" properly it is made to also drive all the contact rolls in the finishing leg.

The take-up end capstan at the exit of the plating line, in conjunction with the take-up spool is fundamentally the tension establishing means for the entire plating line.

From the take-up capstan, the plated wire is passed back to the main floor over another fanning section, and through guide tubes to the take-up spools. While the take-up capstan may be said to establish tensions for all of the plating line up to it, it is the take-up spool drive which basically originates and determines the order of wire tensions for the entire machine. Any tension increase or decrease originating at the take-up spool is reflected all the way back through the machine to the magnetic supply capstan (Fig. 4).

On the electroforming machines wire tension control is quite important. If tensions become too low, the wire may intermittently lose contact on the rolls, sag or weave in the plating cells enough to disturb the spacing between the wire cathode and the anode bed in the cell, or actually stall or hesitate long enough to be burned in two at a contact roll. On the other hand, if tensions are permitted to become too high, the tight wire produces excessive wear on the contact rolls, and bruises or scrapes the relatively soft thin lead and brass plate on the wire. For these reasons the wire tensions developed by the take-up spool drives must be controlled within narrow limits.

The power required to drive a take-up spool is practically constant from an empty to a full spool, but the rotary speed of the spool must decrease as the spool fills up, while the driving torque must increase to compensate for the increase in winding radius. The take-up spools are therefore individually driven with compound-wound dc motors which automatically slow down as the spool fills up, meanwhile holding wire tensions between 15 and 20 pounds from start to finish of a spool.

All the plated wire produced from a 450-pound core wire supply is taken up on one spool to eliminate intermediate wire cuts, and to produce



Fig. 4 — Finished wire, after a series of cleaning and plating operations in the electroforming process, is wound onto take-up reels for final inspection and shipment to the Wire Shop.

the longest lengths of plated wire that can be made from a single reel of core wire. This means that for each core wire reel that goes onto the pay off stands, one spool of plated wire comes off at the take ups. This practice reduces scrap operating losses and spool handling, and simplifies the keeping of production records. The take-up spools, which nominally hold 600 pounds of plated wire, were made large enough to take 750 pounds. This allows steel wire suppliers up to 50 pounds overrun above 450 pounds of core-wire per spool.

The spools are made with large arbor-holes which are slightly countersunk to provide seating surfaces for engaging cone plates which hold the spool in the take-up stand. The cone plates clamp the spool between them under the pressure of a direct-acting compressed air cylinder closure, which permits quick loading and ejection of the spools and greatly reduces the time required to make spool changes thus keeping scrap to a minimum. Driven pinch rolls are provided at each take-up stand to draw the wire from the take-up capstan while a spool change is being made. The pinch rolls do not contact the wire while it is being wound onto a spool. Operating linkages are provided whereby the pinch rolls are applied to engage the wire at the instant the take-up spool is stopped. In this way there is no pause and no interruption of the wire's motion from the take-up capstan. Wire passing through the pinch rolls is slightly flattened, so it must be scrapped. Spool changes can be made quickly enough so that no more than 10 feet of wire in 125,000 feet have to be scrapped from this cause.

The wires are laid evenly on the spools with a gang-distributor driven by a separate motor. The spool traverse is wide, the wire lay is close, and the distributor travel can be accurately set, both at the distributor traversing screw and at the individual take-up position, so that the wire will be distributed on the spools without camber or end pile-up. Every precaution is taken to insure good wire distribution on the take-up spools, because the wire is subsequently taken off them at high speed at the insulators where wire breaks resulting from faulty wire distribution cannot be tolerated.

PLATING MACHINE

The plating machine frame is a simple structure built up of regular structural steel sections and lined with sheet steel. The cross section of the machine resembles the letter "H" with the cross bar set low. The tops of the "H" are joined to lengthwise channels, which form a continuous rail. The bottoms of the "H" are welded to heavy longitudinal channels which transmit the weight of the machine through rollers to a

pair of heavy lengthwise "I" beams which form a part of the building structure. Supported within this framework is the "U" shape trough of stainless steel or low carbon sheet steel covered with Koroseal lining if required by the electrolyte utilized. A step is formed in the bottom corners of the "U" to support the plating cells leaving a trough-shaped channel in the center for the return of the electrolyte. The header from which the electrolyte is supplied to the individual plating cells is located between the legs of the "H," flanked on one side by the positive electrical bus and on the other by the negative bus. Partitions are provided as required in the length of the trough to separate sections in which different electrolytes are confined. The entire machine is sloped towards the central portion of the building so that the effluent electrolyte is directed to the low portion where a downspout to the solution storage tank on the floor below is provided.

The actual plating operations are carried out in a series of cells which are supported on the steps in the return trough by insulating blocks (Fig. 5). All cells are of the same basic design. A typical cell consists of a "U" shaped body of formed and welded sheet steel which may be low carbon steel, stainless steel or low carbon steel Koroseal lined and covered, depending on the particular electrolyte involved. Studs are mounted in both ends of the body for fastening the weir plates. These are of molded hard rubber and provided with twenty-five equally spaced slots on 1-1/2" centers to pass the wires. The weir plates on the two ends of the cell body differ in thickness, the thick weir containing an interior manifold which serves to distribute electrolyte uniformly across the cell and which is connected to the electrolyte supply pipe by a flexible rubber ell. Molded rubber spill catchers are bolted to the outside of the weir plates to collect the electrolyte discharged from the weir slots and direct it through a rubber tube into the return trough with a minimum of splash.

In the case of unlined cells, the cell body serves as anode or cathode as the operation may require and the electrical connection from the bus bar is made directly to a tap on the cell body. For the lined cells, an anode plate of a suitable metal is provided in the bottom of the cell and covered with anode material in the form of cast shot. In this arrangement, all the electrolytic corrosion takes place on the shot bed, leaving the lead-in plate undisturbed.

The thick, or feed weir, is placed on the low end of the cell to take advantage of friction loss in the long weir slots to reduce the discharge of electrolyte from the cell. Maple wedges coated with vinyl paint placed between the cell bodies and the trough sides allow easy and accurate

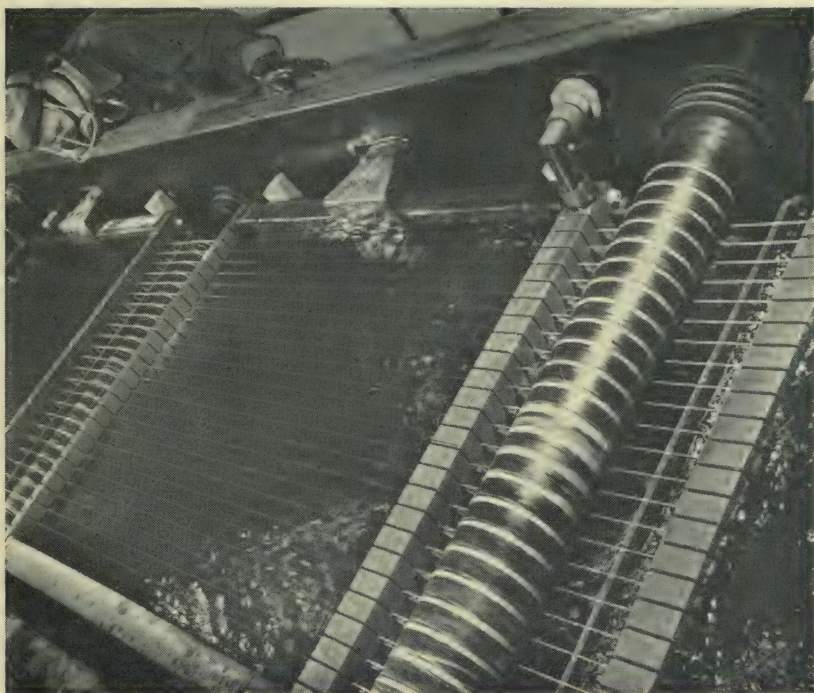


Fig. 5 — Typical cell of the many in each of the two 600-ft. long electroforming machines shows how the twenty-five channels of wire pass through a plating solution and are propelled and fed electric current by rolls like that seen in the foreground. An operator, wearing his protecting safety gear, checks over the operation.

alignment of the weir plates with the wires and maintain the cells rigidly positioned.

The electrical circuit for the plating current is completed through the contact rolls which position and make contact with the wires between plating cells. The rolls are heavy-walled tubes carried on axial shafts by internal ball bearings and insulated from the machine frame. The materials of construction are copper, steel, stainless steel, or monel metal, depending on the electrolyte involved. The ends of the tubes project through the trough walls and the shafts are carried on external brackets which bolt to the upper longitudinal frame channels. One of the brackets is cast iron and the tube on this end is furnished with an insulated sprocket for the driving chain. The other bracket is a copper casting, insulated from the machine frame and carrying brush holders for copper-graphite brushes which contact the opposite end of the tube. The plating

current is carried from the copper casting by a flexible connection to the bus bar. The surface of the tube is provided with shallow grooves on 1- $\frac{1}{2}$ " centers to position the wires and the whole contact roll assembly consisting of tube, bearings and shaft can be shifted laterally to bring a new set of grooves into play when wear dictates.

Pilot plant experience had shown that electrolyte will be carried by the wires to the contact rolls and will spread out over their surface, eventually reaching the bearings and the brush contacting surfaces. Some type of seal had to be provided where the roll passed through the wall of the trough. Various types of commercial seals were tried, but all left much to be desired. The seal finally developed for the application is in two parts: a molded Neoprene double slinger ring on the roll, operating in a molded Neoprene housing inserted in the trough wall. The slinger ring successfully prevents the passage of electrolyte while the housing prevents splash from carrying past. This seal has the important further advantage that there is no contact between the fixed and moving parts so that there is no friction loss and no wear. Also, it is a simple molded rubber part which slips into place and requires no fastening.

The wires, in their travel through the plating machine, are acted upon by nine different electrolytes. An appreciable amount of electrolyte is carried with them and if means were not provided to remove the envelope of solution, the succeeding baths would be quickly contaminated, and in some cases dangerous gases would be generated. In general, the transition section between dissimilar electrolytes is made up of an air wiper, a water wash and a steam wiper, in order of wire travel. The wiping element at each wire is the convergent blast of steam or air from three small nozzles. Twenty-five sets of these nozzles, one set for each wire channel, are mounted in a manifold. In the case of an air wiper, the spent air and droplets of electrolyte wiped from the wire are directed into an eliminator where the droplets are caused to separate from the air stream by impinging on the metal baffle plates. The steam wiper is of similar construction except that a water-cooled condenser is substituted for the eliminator. The cooling water for the condenser is discharged into a typical cell, mounted between the two wipers where it washes the wire and discharges to waste.

While these washing facilities may appear somewhat elaborate, they are justified on the basis of reliability and safety. The failure of any one of the three services, air, water and steam, will not cause appreciable contamination before repairs can be made.

At the copper cyanide, brass cyanide and acid copper plating sections, the wipers are preceded by dragout recovery units. A dragout recovery

unit consists of a cell and a small, individual reservoir with a small pump which circulates the liquid in the reservoir through the cell, the overflow returning to the reservoir. Makeup water to replace that lost by evaporation in the associated plating section is added to the reservoir, serving to continuously dilute the electrolyte washed from the wire. The reservoir operates at constant level, the excess of dilute electrolyte passing on as makeup. The dragout recovery units act as an additional safeguard to limit the loss of valuable plating solution and to minimize the contamination of wash water going to sewer.

Last in the design of the machine are the precautions taken to protect against stray currents and to provide for expansion and contraction. Protection against stray currents is an important consideration in any equipment employing highly conductive liquids and heavy currents. All plating cells are insulated from the machine frame by their supporting blocks. Both positive and negative plating circuit buses are insulated. All electrolyte pipe lines contain neoprene flexible joints which serve both to sectionalize them electrically and to provide for expansion and contraction. Contact rolls are insulated at their mountings. These are the principal precautions taken.

The expansion and contraction problem was rather complex. The three trough sections of each machine are continuous units, the shortest 77 feet, the longest 300 feet in length. The plating cells in the trough are supplied with current and electrolyte from a number of different locations in the building and the building itself is provided with a single expansion joint in the center. To allow for relative motion between machine and building, the machine sections are mounted on rollers. Since these sections are required to operate at various temperatures, from room conditions to 195°F, and are required to be supplied at frequent intervals with current and electrolyte, all electrical connections to bus bars are made with flexible joints and electrolyte is supplied to each plating cell through a special flexible molded ell.

Each machine section is anchored to the building steel to prevent creep. An analysis was made of the several movements to be expected under various operating conditions and ambient temperatures and the anchor point was selected for each section so that the relative motions are minimized.

PLATING OPERATIONS

From the chemical point of view, the electroforming process consists of a series of unit processes in tandem, to clean the steel core wire and to successively deposit the several metallic coatings required to make

the completed conductor. In the course of the complete travel thru the machine the wire receives thirty-two separate treatments in nine different chemical solutions.

The first solution that the wire enters is the alkali cleaner which removes oil, drawing compound and dirt. The heated bath contains an alkaline solution with a small quantity of a wetting agent. This section of the machine contains eight stainless steel cells alternated with seven stainless steel contact rolls mounted in a stainless steel trough. The wire is anodic with the body of the cells acting as the cathode. A current density of 100 amps./sq. ft. is applied to the wires causing a heavy ebullition of gas which materially helps the cleaning operation. The alkali cleaner section is followed by a steam wiper.

Next the wire passes into a sulfuric acid pickle section where scale, rust and occlusions are removed and a slight etch is imparted to the surface of the steel to promote adhesion with the subsequent copper deposit. A small quantity of an inhibitor is added to prevent the dissolving of an excess amount of iron which would result in a heavy carbon smut on the surface of the conductor. There are six cells and three monel metal contact rolls. Following the pickle is an air wiper, a water wash cell and a steam wiper. This completes the cleaning and preparation of the wire surface prior to the first plating operation. The tank and cells in the sulfuric acid are constructed of Koroseal lined mild steel.

The initial coating is a thin layer of copper from a copper cyanide solution termed a "cyanide flash" and is designed to give a smooth deposit. There are five plating cells and four contact rolls of low carbon steel and the machine trough is likewise of low carbon steel. The wire is cathodic and the copper is deposited at a relatively low current density.

Following the copper cyanide flash the wire passes directly to the copper cyanide plate solution where a copper coating of not less than 0.0001" thickness is applied in a bath designed to operate at a high current density and, therefore, at a higher rate of deposition than is obtained in the flash bath. Seventeen plating cells alternated with sixteen contact rolls are needed to deposit the required thickness at the operating current density. Cells, rolls and troughs are made of low carbon steel, and copper shot resting directly on the steel cell bottoms forms the anode surface. The cyanide plate section is followed by a dragout recovery unit, a water wash cell and a steam wiper. This completes the preparation leg of the machine.

When the wire leaves the preparation leg, it passes through a turning section (Fig. 6) which reverses the direction of travel prior to entering the acid copper plate leg where the bulk of the copper is deposited.

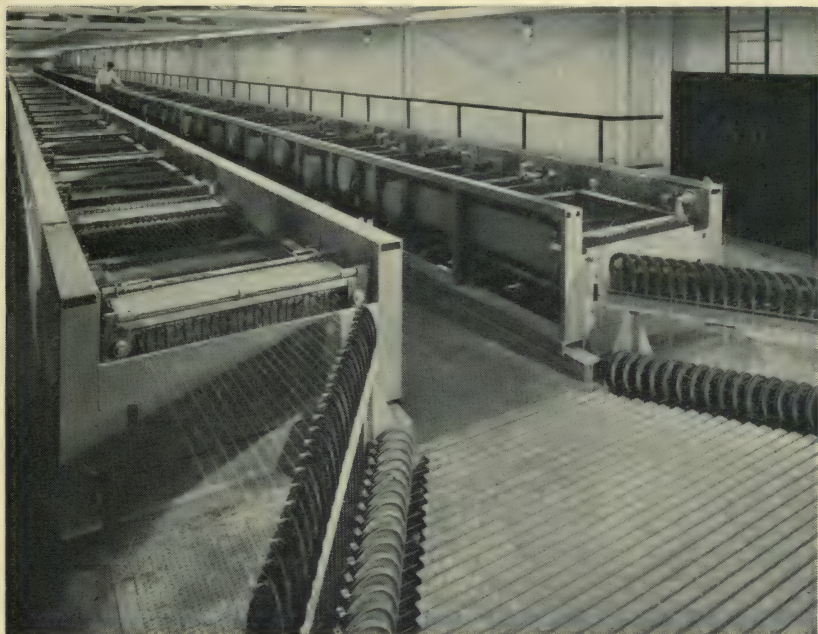


Fig. 6 — Turning section of the machine where the twenty-five channels of wire, after having been cleaned and partially plated in the tanks on the left, reverse their direction and travel the full length of the building, receiving added copper coating in the series of plating cells.

The production plating machine has 58 plating cells alternating with 57 copper contact rolls. The plating solution is very corrosive so that all surfaces of containers are covered with Koroseal. Copper plates in the bottoms of the cells distribute current to a bed of copper shot which forms the active anode surface.

A relatively large number of cells is required because of the magnitudes of the total plating currents involved. Faraday's Law shows that several thousand amperes are required to deposit copper at the rate of 0.001" per minute.

Instantaneous fusion of a steel wire 0.033" in diameter would result if this current were forced through it at one time. The repeated passage of smaller currents which will not overheat the wire will, however, deposit the same amount of copper. The design of a plating section thus takes the form of a number of plating cells separated by contact rolls.

The contact rolls are designed to have a wall thickness of copper adequate to prevent more than a negligible voltage drop between the two outside wires in the machine. This insures that the same voltage is ap-

plied to all wires across the machine and permits a design in which the current collecting brushes are always located on the same side of the machine.

When the wires leave acid copper plate cell No. 58, the full deposit of copper has been applied. They pass into a dragout recovery cell where they are washed by the makeup water, through an air wiper, and a water wash cell before entering the heat treat section. The heat treatment has two functions: First is to change the grain structure of the deposited copper to an annealed form having small random crystals free of strain. Second is to strain relieve the hard drawn steel core wire sufficiently to increase its elongation to between 3 and 7 per cent. This is accomplished by passing current through each wire to heat it to the necessary temperature. The heat treated wire passes through a water wash cell which serves as a quench and through a steam wiper which dries it in preparation for entering another turning section where it again changes direction 180° to enter the finish leg of the machine.

The first solution that the wire enters at the finish leg is hydrofluosilicic acid which serves to remove any oxide which may have formed in the heat treatment. This cleaning operation is performed in a single koroseal covered cell.

Four koroseal covered cells, three copper contact rolls and a koroseal lined section of trough make up the lead plate section. The electrolyte is lead fluosilicate. Lead sheets in the bottom of the cells are covered with lead shot to form the active anode surface.

The brass plating section applies the final deposit to the wire. Its function is to provide a coating which will unite chemically with the insulating compound, giving good adhesion so that the load from the drop wire clamps used to support the wire in service will not cause the insulation to slip on the conductors. The composition of the deposited brass is controlled between very close limits to obtain the desired adhesion between conductor and insulating compound. The electrolyte contains copper and zinc cyanide. There are four steel plating cells, three steel contact rolls and a low carbon steel trough in the brass plate section. The anode material is a mixture of copper and brass punchings which rest directly on the steel cell bottoms. The finished wire is then wound onto 660-pound reels.

ELECTRICAL POWER

Electrical power is purchased from the local utility company at 13.2 kv and brought into the building by a pair of three-conductor, 300,000 circular mil, 15,000-volt, lead-covered cables running in underground

ducts. These are terminated in two 1500-kva metal-clad substations which consist of high-voltage switching units, transforming units, and low-voltage feeder units. Each unit substation has the self-cooled rating given in Table I.

On the incoming line side of each substation are two load interrupter disconnecting switches so that the building may be supplied by either of the two incoming high-voltage cables. The high-voltage switches feed the power into the two transformers, each of which is rated at 1500 kva, three-phase, 60-cycles, 55°C rise, and is self-cooled, filled with non-inflammable liquid. For a short time interval, with proper switching, either of these transformers can supply the total building load if fan cooling is provided.

The 480-volt power from each transformer is then fed into its respective low-voltage switchgear for distribution to various parts of the build-

TABLE I

Capacity (55°C.)	1500 kva
Normal voltage	13200/480-277 delta-wye volts
Frequency	60 cycles
Phases	Three
Circuits:	
Three-wire incoming, 12300 volts	Two
East unit, four-wire, outgoing 480/277	Six
West Unit, four-wire, outgoing 489/277	Five

ing. Lighting, heating and ventilation services are provided totally from one substation by way of a disconnecting bus tie switch which automatically throws over to the other substation in case of failure of one.

Other than the lighting, heating, ventilation and some plant services, each substation is designed to supply power for the operation of one machine.

From each substation one low-voltage ac feeder supplies on its machine a number of rectifiers which transform and rectify the 480-volt, three-phase, ac power to dc power at a low voltage. From this feeder is supplied all of the "auxiliary" plating power for that machine: alkali cleaner, acid pickle, cyanide copper flash and plate, electrocleaner, lead plate and brass plate. In addition the dc power for establishing the field in the magnetic capstan is supplied from a rectifier on this feeder at 55 volts.

Another low-voltage ac feeder from each substation feeds through an induction voltage regulator which in turn feeds sixteen rectifiers of one machine. From these rectifiers is taken the dc plating power for the acid copper plate section of the machine. The function of the induction voltage

regulator is to automatically control the dc plating current in the acid copper plate section through a control system which continuously measures the electrical resistance of the product and signals the need for variation in power in accordance with the variation in resistance.

All of the rectifiers associated with one machine (Fig. 7), including the "auxiliary" units and those on the regulated feeder, are located in the outer mezzanine bay nearby that machine. The two regulators, although each is associated with a different machine, are both located in the sub-station room on the south side of the building.

The induction voltage regulators are basically standard regulators of the type used in lighting service but modified slightly by the addition of a control slidewire for electrically indicating the position of the rotor. Each unit is three-phase, self-cooled by non-inflammable liquid, rated at

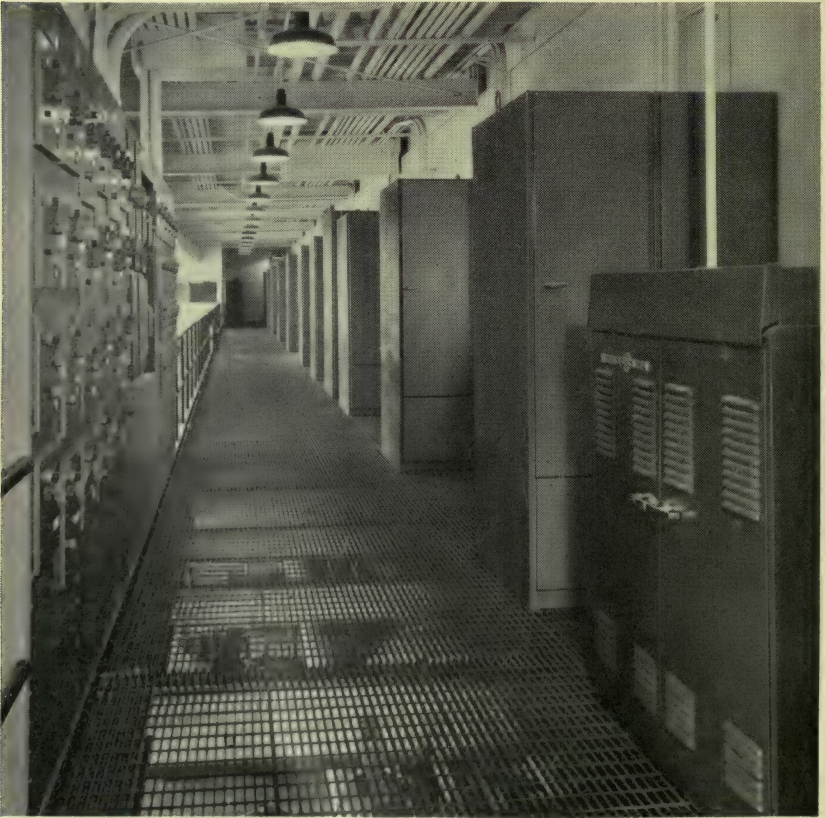


Fig. 7 — Electrical heart of the electroforming process are these rectifiers and controls where alternating current is converted to direct current.

480 volts ac and having at least 105-kva capacity according to NEMA standards. Each regulator is able to buck or boost line voltage by 20 per cent at maximum rectifier load.

The rectifiers used for converting three-phase, 60-cycle, 480-volt ac power to low voltage dc are all of the copper-oxide "dry-disc" type. A detailed investigation of various sources of dc power was made before this type of converting equipment was purchased. Because of the similarity between the continuous copper plating of wire and the continuous tin-plating of sheet steel, a tour of several steel plants which plated strip steel in a continuous process was made. Here were observed some of the best and some of the poorer installations of electrical converting equipment, including motor-generator sets, copper-oxide rectifiers and selenium rectifiers. This survey terminated in a decision to use totally enclosed, air cooled copper-oxide rectifiers, in which the enclosed air is recirculated past air-to-water heat exchangers for cooling purposes. This decision was based on a desire for low maintenance costs, efficiency, compactness of installation and flexibility. Choice of copper oxide over selenium rested on the fact that copper oxide rectifiers are less costly and better suited for low voltage output (less than 6 volts).

The actual design of the rectifiers was arrived at after a series of conferences between Western's engineers and those of the manufacturers. Accessibility of all parts for inspection and maintenance purposes, adequate protection of components through control circuits, limited occupancy of floor space and low cost were design criteria. These features are nicely combined in the final design and the manufacturer standardized the rectifier for sale to other purchasers.

Each rectifier is of the completely enclosed, air recirculating type. It is equipped with its own heat exchanger and is gasketed to minimize air leakage. The air-to-water heat exchanger is of a type especially designed for use with well water containing some fine sand. It features ease of cleaning and is a type suitable to these conditions. The fins and tubes are of copper and water connections are external to the rectifier housing.

All other components are readily accessible through gasketed doors in the housing. One or two fans, as required, circulate the air within the rectifier housing to transfer the heat generated in the rectifier cells and in the transformers to the water cooled heat exchanger. As protection both to equipment and to personnel, a rectifier can be automatically tripped off the line for any of the following reasons: Electrical overload and short circuit, dc over-voltage, failure of air circulation, overload of fan motor, water failure, momentary power interruption, over-temperature of the air, and by opening any of the doors on the rectifier.

Control of the output of the entire bank of rectifiers of the acid copper plate section is accomplished through the associated induction voltage regulator. Individual adjustment of load and compensation for aging of the rectifier stacks has been provided for by bringing out to terminal boards within the housing a number of leads from the primary windings of the rectifier transformers. For a given feeding voltage, the output of each rectifier can be adjusted over a considerable range by the positioning of jumpers at the terminal boards. Except for aging compensation, this adjustment, once made, need not be made again. Control of the output of the "auxiliary" rectifiers is accomplished by tap-switching, under load, between taps brought out from open-delta autotransformers on the primary side of the rectifier transformers.

The capacity of the bank of the sixteen rectifiers feeding the acid copper plate section of each machine is larger than is actually required to produce the wire. With this additional capacity distributed among the rectifiers, it is possible to continue plating operations in case of failure of one rectifier. The automatic control system will increase the output of the remaining fifteen units immediately after the one has dropped off the line.

The bay in which all rectifiers are located is separated from the operating bay by a plaster partition which serves the purpose of forcing the incoming fresh air flow past the rectifiers protecting them from any corrosive atmosphere which might be present in the operating bay. The rectifiers are placed against the wall and their bus terminals extend through small apertures into the operating bay. Because the operating bay floor level is about two feet higher than that of the rectifier bay, the rectifier terminals emerge from the plaster partition below the floor elevation of the machine. This permits the running of the busbars out to the machine by passing them under the operators' walkways.

These feeder busbars from the rectifiers terminate beneath the section of the machine which they are to feed. Flexible laminated connectors join the feeder buses with the buses running longitudinally under the machine: from the longitudinal buses additional flexible connectors carry the current up to the rigid terminals leading into the plating cells or up to the copper castings supporting the brushes and contact rolls.

The dc circuit, both inside and outside the rectifiers, has been completely insulated from ground excepting that grounding caused by electrical contact with the wire itself. Stray currents are thereby minimized and kept out of pipe lines or building steel.

All of the busbars used on the machine are of the square drawn copper tubing type. Aluminum was considered but rejected because of the seri-

ous danger of electrolytic corrosion at the joints (which necessarily had to be made to copper) from the warm, moist, corrosive atmosphere. The square copper tubing is ventilated by holes drilled in the upper and lower faces to give it comparatively high current carrying capacity with low temperature rise. The fact that the bus is in the form of tubing gives it an inherently high mechanical strength; the supports for the tubing, therefore, could be spaced much farther apart than for ordinary bus. All tubing bus joints are of the compression type made up by the use of standard clamps rather than by brazing or bolting. All of the tubing was purchased cut to size and coded. During installation, the preparation of the joining surfaces was carried out on benches in the area. Such planning resulted in an exceptionally easy and rapid installation.

A complete picture of the electrical operation of a given machine is provided to the operator through a control console located in the operating bay of the mezzanine (Fig. 8). On the control console are located the tapswitches for adjusting the output of the "auxiliary" rectifiers, an ammeter and a voltmeter for each rectifier (a total of forty-six such meters per machine), a meter which indicates wire speed, switches for

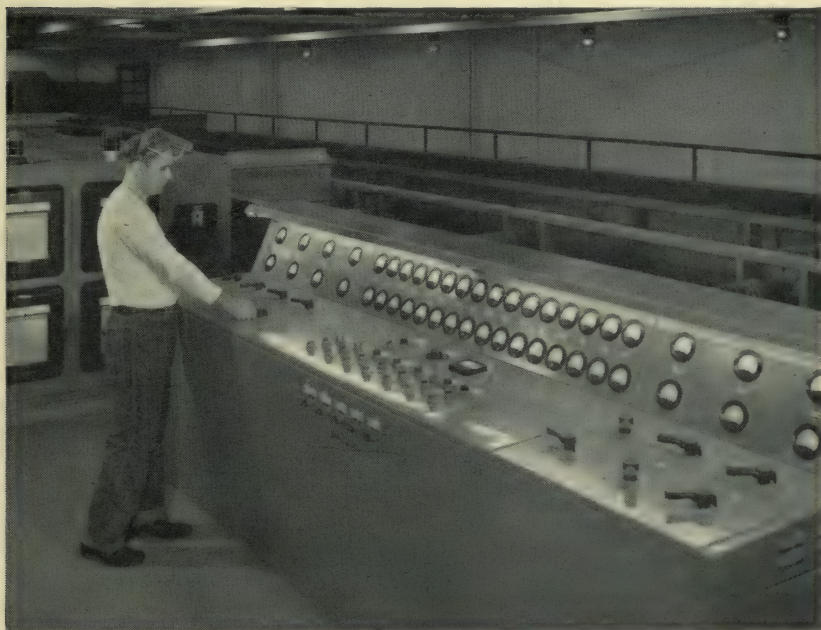


Fig. 8 — Control console for the entire process is readily available when necessary.

the ventilating roof fans, a switch for removing the magnetic field from the supply-end capstan, a selector switch for selecting the speed of wire passage through the machine, and seven push-button stations which give the operator individual control over various elements of the machine. The push buttons energize relays mounted in a bank within the control console and thus permit a large degree of interlocking. The interlocking is designed to prevent accidents from occurring in the process. For example, the operator cannot permit the pickling acid to flow over the wire while the wire is stationary; otherwise the acid would dissolve the wire in a short time. Time delay relays also are used to advantage. They insure, for example, that the full length of moving steel core wire has its flash coating of cyanide copper prior to the rise of the acid copper plate solution into the cells and around the wires which, without the copper envelope, would be eaten through in a short time.

The overall machine control system has been broken down into two parts because of the two-floor design. From his console, the second floor operator controls only those portions of the process which are visible to him on the second floor. He can, for example, permit solution to flow into the plating cells by electrically opening the valves in the circulating system. Actual control of the first-floor pumps which move the solution, however, is in the hands of the first-floor operator who has a control panel strategically located near each storage tank and pumping station. Here he can operate filters, pumps and level control systems.

WIRE RESISTANCE CONTROL

There has been provided for each machine in the electroforming plant a control system which automatically adjusts the dc plating current in the acid copper plate section to that value which, under all normal operating conditions, will produce a wire meeting the electrical resistance specification for the product.

Control is necessary because of the number of variable factors which, if uncontrolled, would interfere with the proper flow of plating current and, therefore, with the amount of copper deposited. These variable factors can occur in the physical and chemical properties of the raw materials entering the process, in the plant services feeding the process, and in the process itself. Some of the more important variables are as follows:

1. Incoming electrical line voltage fluctuations.
2. Failure of one of the sixteen rectifiers feeding the acid copper plate section.

3. Changes in the temperature or in the concentration of the electrolyte.

4. Changes in the distance between the anode bed and the wire because of the wasting away of the anode bed during plating operations.

5. Wire speed changes (by removal of the magnetic field from the supply-end capstan).

6. Operation of the machine at less than its capacity number of channels.

Control is justified by the savings in copper which accrue, by the relieving of the operating personnel of tasks which could become tedious, and by the resulting production of an exceptionally uniform product. In view of the size tolerances available on steel wire, it was decided that the most practicable approved way was to manufacture to uniform conductivity and to ignore wire size within reasonable limits.

Two systems produce the overall control function. A primary system senses the total value of the dc current in the acid copper plate section by means of a current transformer located in one phase of the common ac feeder servicing all sixteen rectifiers. Changes in that current are immediately detected and compensated for by adjustment of an induction voltage regulator located at the head of the feeder. This system therefore maintains the dc plating current essentially constant in value and corrects for such fluctuations as are listed from 1 to 4 above before they can seriously affect the product.

Changes in certain other variables, as exemplified by 5 and 6 of the above list, require that a new value of total current be established and maintained constant by the primary system. Otherwise the maintaining of the previous value of total plating current would result in a change in the resistance of the product. Such variables are cared for in a secondary control system which uses the resistance of one of the finished wires as the control parameter. The secondary system automatically positions the control point of the primary system, depending on the value of electrical resistance of the pilot wire which is measured continuously.

In addition to control equipment each machine has an inspection device which automatically measures the resistance of each of the twenty-five wires and records the values on a chart. A cycle of all twenty-five wires is completed each hour.

Both the control and inspection resistance measuring equipments use Kelvin Bridge circuits and measure approximately five feet of wire at one time. Continuous contact with the wire being measured is made through rotating sheaves and brushes. This avoids any scraping of the soft lead and brass surfaces of the product.

The inspection wire contacting device automatically moves from wire to wire across the machine, sampling the resistance of each wire for about fifteen seconds. The chart provides the observer with a complete picture of the behavior of the plating operations at a glance and enables the operator to properly select the controlling pilot wire in such manner that minimum energy and copper are expended. The control wire contacting device is manually set in position against one wire until conditions require selection of a new pilot wire.

Both contacting devices are located in the position just preceding the take-up capstan from which the wire moves down to the first floor to be taken up on reels. To avoid the by-passing of the bridge current through ground and the building steel, each channel has been completely insulated from ground and from other channels between the contacting devices and the end of the wire on the take-up spool. The take-up capstan, all sheaves, guide tubes, rollers, and the take-up stands themselves have been designed to provide the required insulation.

Aside from presenting to the operating personnel a complete chart showing the values of the electrical resistance for each channel, the sensitive resistance measuring equipment also gives indication of process difficulties long before they are evident from any other source. For example, an increase of organic contamination above certain concentrations in the acid type plating bath will produce wide variations in the structure and, hence, in the electrical resistance of the copper deposit. These fluctuations can occur in very short sections of the product. While the average value of the resistance may remain constant for long sections of the product, the resistance charts provided by the control and inspection equipment will record the individual wide fluctuations and thereby provide experienced operating personnel with a positive indication of imminent serious trouble.

SOLUTION HANDLING

In this plant, all unit processes are similar if not identical in arrangement and operation.

A typical handling system consists of a storage reservoir, mixing and filter station, circulating pumps, heat exchanger, solution supply, and return piping between storage tank and processing area, and a solution temperature and level control. (Fig. 9.) The chemicals of which the electrolyte is composed are placed in the mix tank and agitated. This mixture is then transferred through the filter into the storage tank. The filtered solution in the storage tank is first circulated through the heat exchanger and finally delivered to the processing area on the mezzanine

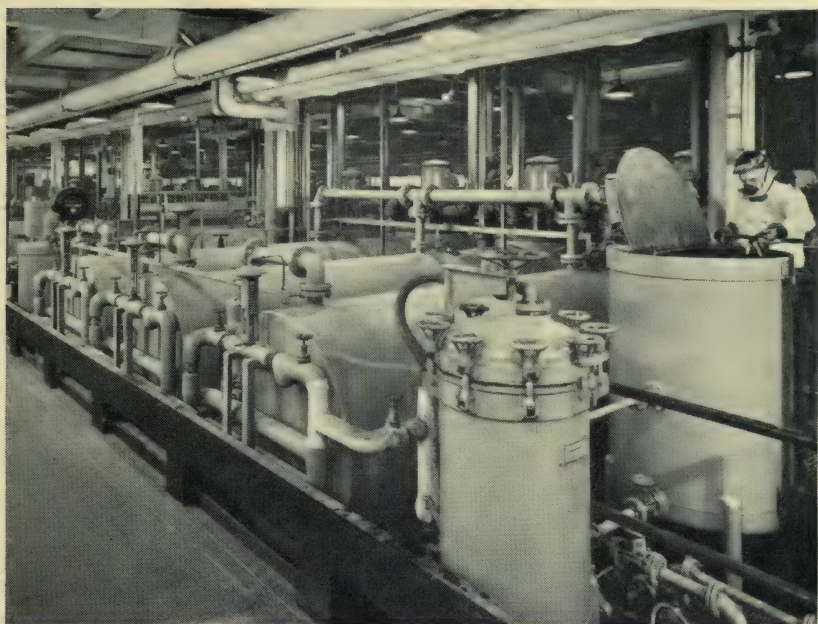


Fig. 9 — Solutions for the plating operations are stored in this battery of tanks. Full safety equipment is worn by the operator as he performs the job of tank-filling.

floor. It returns by gravity to the storage reservoir and the cycle is repeated.

The storage reservoirs are cylindrical horizontally mounted steel tanks. Those containing acid solutions are Koroseal lined, and many are covered externally with thermal insulation. All tanks have the same inside diameter, but vary in length depending on storage requirements for each electrolyte. Since the electrolyte used in similar unit processes of the two machines is the same, provision has been made for complete interchangeability of solution between the storage tanks of the two machines.

The purpose of the filtering operation is to permit the continuous removal of suspended solids and foreign material in the process solutions. The mixing and filter station provided for this purpose is an assembly of components all of which are mounted on a common base or platform. The basic assembly consists of a closed pressure type alluvial-leaf filter, centrifugal pump, mixing tank with agitator, various valves for directing and controlling the quantity of solution, water and air used in the filter operation, and a discharge to sewer for cleaning. Within the pressure tank a series of double faced screen assemblies or leaves comprise the

filtering elements which hold back the undesirable solids and discharge a clear product into the common filtrate manifold on which they are mounted. There are eight leaves in each pressure tank providing sixty square feet of filter area. The filter cycle is begun by filling the mix tank with unfiltered solution or electrolyte from the storage tank. Sufficient filteraid for precoat is rapidly mixed in by agitator. This mixture is forced through the filter leaving a deposit or coat of filteraid on the intake surface of the leaves. The purpose of the coating is to provide a filter medium which is dense enough to prevent the passage of suspended solids in the solution and yet is easily removed and replaced. The appearance of the liquid on returning to the mix tank gives an indication when an adequate precoat has been deposited on the leaves. Once the coat has been applied, the filter is put on the line and clear filtrate is discharged to the storage tank. The filtration stage terminates when the filter tank pressure approaches a maximum operating value, at which time the leaves are washed, a new coat is applied, and the cycle is repeated.

The storage reservoir is fitted with one or more heavy duty vertical sump type pumps to transfer the electrolyte in the processing area. All pump parts in contact with chemical solutions are constructed of either cast iron or a stainless steel alloy known in the industry as FA-20. All pumps are fitted with graphitar bearings. Each pump is connected directly to its electric motor and the entire assembly may be quickly removed from the storage tank.

All piping associated with solution handling is either steel or saran-lined steel. The Saran-lined pipe has flanged type joints and the unlined steel pipe has welded joints except where disassembly is provided for and here welded-on flanges with neoprene gaskets are installed. Flexible neoprene joints are provided in the solution riser and return pipe to insulate the plating machine from the circulating pumps. To control solution flow to the processing area, air operated diaphragm valves have been installed.

The two-pass heat exchanger selected for the process consists of a steel shell, but all parts in contact with corrosive solutions, including the tube bundle, are constructed of a chemically inert carbonaceous material. The heat transfer rate of this graphite base material is the highest of all non-metals and is higher than that of most metals. For example, the thermal conductivity of 18-8 stainless steel is approximately one-fifth that of the graphite base material.

Operating temperatures for all solutions are specified and equipment to control these temperatures has been provided. The temperature of each solution is measured with bulb type expansion thermometer. The

temperature is recorded by an instrument attached to the outer trough of each process on the mezzanine floor. In addition to recording temperatures, the instrument also operates a spring loaded valve, by air pressure, in the heat exchanger piping. The correct operating temperature can be chosen by setting the manually operated pointer on the recording chart. Automatic, electrically operated, liquid level control devices have been provided for each solution. These controls are provided for the following reasons:

1. The evaporation of water from the solutions used in the process is of such magnitude that some means of maintaining solutions within their proper concentration range is necessary.

2. The solution level in the storage tank must not reach a specified lower limit, or the circulating pump suction head will fall below the minimum operating value.

3. When the electroforming machine is in operation, the total solution volume in any unit process is composed of the portion in storage and the portion in circulation. If the portion of solution in the storage tank exceeds a known value, the remaining volume of storage space will be insufficient to accommodate the circulating portion of solution which returns to storage when the machine is shut down.

The liquid level control device consists of four electrodes which dip into the contained solution. The four electrodes are arranged so that as the level varies, the solution closes different electric circuits. During normal operation at proper level, the electrical connection between the low electrode and a common electrode holds the solenoid operated water valve closed. When the level falls about $\frac{1}{2}$ " below the prescribed level, the circuit opens and releases the solenoid water valve to the dragout recovery cell or the valve in the water line leading to the filling nozzle in the storage tank. When solution again rises to the prescribed operating level, another tripping circuit is activated which causes the water valves to close.

In order to show whether or not the level controls are operating satisfactorily, a red and a green light are mounted above and in line with each storage tank. When the green light is on, the levels are proper. The red light is a signal that the level is increasing beyond the point at which the tripping circuit should have been activated. In this case, the water valve is immediately turned off. When both lights are on, the solution level in the reservoir is low and make-up water is added automatically.

Due to the chemical action of fluoboric acid on the copper anode material in the acid copper plating bath, the copper concentration of the solution rises during operation of the electroforming machine. In order

to hold the concentration within the established operating range, copper removal equipment was developed.

The equipment consists of two Koroseal lined tanks installed in a concrete pit and fitted with graphite anodes and sheet copper cathodes. When the copper concentration of this solution exceeds the maximum operating value, the pump associated with the copper removal unit automatically cuts into the acid copper plate equalizer line and transfers solution to the removal unit. The excess copper in solution is electrolytically deposited onto copper cathodes, and the restored solution is returned by gravity to the equalizer line.

RAW MATERIALS HANDLING

Except for the anode material (shot), all materials handling is confined to the main floor. In this way the plating deck operators are fully relieved of all materials handling chores, other than the one job of replenishing anodes in the cells, and this job is purposely delegated to them because anode maintenance is a critical part of the plating deck operator's responsibility for product quality. All heavy, bulky units are handled on the main floor so that none of the untidiness and confusion always attending the opening of containers, truck movement, and receiving and shipping operations are present on the plating deck to interfere with the prime job done in that area.

All chemicals to be used in the various cleaning and plating solutions are stored along the north wall of the building, each in the area nearest the preparation and mixing equipment in which it will ultimately be used. Wherever chemicals are stored close to one another, which, if accidentally mixed, might create hazards, solid barrier walls are provided to keep them separate. Chemicals are stored in their shipping containers, all of which are fully identified as to contents and adequately marked with any warning labels which might be required. Chemicals, at the receiving dock at one end of the building, are removed from the motor truck, placed on wood pallets, and transported by electric truck to the storage area. The use of pallets for handling and storing the chemicals reduces the total electric truck haul-time and protects the containers from mechanical abuse while in storage. The wood pallets also permit flushing of the floors without soaking the containers, especially those of paper or wood. Facilities are provided for flushing out the emptied container as well as for washing down the floors after accidental spillage or breakage.

Wire is stored along the south wall of the building physically well

isolated from the chemical storage, handling and mixing operations along the north wall. This minimizes exposing the wires to the splash or vapors from the chemicals which would damage the wire. Core wire is the bulk-iest and one of the heaviest items handled. It is stored in an area near both the core wire pay-off stands and the receiving dock, to shorten the truck-hauls and enable the pay-off stand operator to obtain his core wire supply direct from storage without leaving the operating area. The core wire supply reels are narrow and of large diameter so they can readily be handled upright, and rolled along on their head rims. From the upright position, the reels are picked up with a rotatable grapple on a monorail hoist, turned 90 degrees, and placed head down on a waiting transfer car to be moved to the supply or pay-off stands. The transfer car operates on steel rails flush with the floor so that the car can be spotted close alongside the pay-off stands, without danger of being bumped into them (Fig. 1) The bed of the transfer car as well as the beds of the pay-off stands are built up of gravity roller conveyor sections and in loading position both are at the same height. The reel is easily pushed off the car and onto any pay-off stand. Emptied core wire reels are removed with the same equipment and accumulated in the storage area for return to the supplier.

ANODES AND ANODE HANDLING

Soluble anodes for the plating cells are supplied in the form of random-cast shot or pellets and punchings varying in diameter from about $\frac{1}{4}$ " down to $\frac{1}{32}$ ", the percentage of fines being limited by raw material specifications to control the rate of dissolution in the plating electrolytes and to limit the formation of anode muds or sludge. The copper shot is cast from commercial wire-bar copper. The lead shot is made from commercially pure virgin lead, and the brass punchings are obtained from lead-free brass scrap. The brass anode usually contains too much zinc, so the brass bath composition is corrected by adding pure copper shot in with the brass, the proportions being determined by plating bath and plate composition analysis. The shot is laid evenly in a bed about 1" thick over a relatively thin plate electrode covering the entire bottom of the plating cell. Plating potential is supplied to the soluble anode bed through the electrode.

In the cyanide brass and cyanide copper plating cells the anode material is spread directly on the steel bottoms of the cells, steel not being soluble to any consequential extent in the cyanides. Lead electrodes are used in the lead plating cells. The lead and brass anode materials are

used in such small quantities that they are replenished by hand. Copper, however, is plated out of each machine at the rate of 900 pounds per eight-hour shift. Copper shot added to replenish this plate-out from the cells must be spread out evenly over a cell area of more than 500 sq.ft. extending 280 ft. along the machine. When this is done, the average depth of the material added is less than $\frac{1}{10}$ ". As a quality control on the copper plate, the anode-to-cathode spacing in the cells (copper shot layer-to-wire spacing) is not permitted to vary from the specified spacing by more than 25 per cent at any spot on the cell surface and the average must not exceed 15 per cent. Shot must be added at least once every eight hours to meet this requirement, and the addition must obviously be made while the wires are running. It would be impracticable to do this job manually in any reasonable time and with any assurance that the quality of the electroformed copper deposit would not be impaired during the addition.

Copper shot is, therefore, added to the cells with a mechanical distributor which automatically sprinkles the shot into the cells uniformly while the machine is running. It is essentially a hopper mounted on a motor driven carriage which travels on the two upper rails of the acid leg of the machine, the bottom of the hopper being fitted with a rubber surfaced feeder roll and a stripper comb which releases from the hopper a controlled amount of material. The feeder roll is also motor driven and timed relative to the travel speed of the carriage. Both carriage and feeder roll are operated from rubber grips so that the machine operator may stop the distributor wherever he wishes or by-pass any cell if he so desires. To prevent the spillage of shot on the contact rolls between the cells, block-out cams are placed on one rail to operate a cut-out switch which stops the feeder roll while the distributor passes over the contact roll space between cells. The speed of the feeder roll can be varied, and the height of the stripper comb can be changed to vary the amount of shot released per foot of distributor travel. A feed run of the distributor requires an operator's attendance and takes about twelve minutes to discharge shot to the entire acid copper plating line. The distributor returns to its starting point automatically when the feed-run has been completed. The feeder roll is inoperative while the distributor is returning. Power is supplied to the distributor through trolley operating in fully guarded feed rail running overhead and parallel to the plating line.

Copper shot is added to the distributor hopper with tilting bucket on a monorail hoist. The bucket is filled in the raw material storage area on the main floor, raised with an electric hoist to the plating deck and moved on monorail track to dumping position over the distributor hopper.

FINISHED WIRE HANDLING

The spools of plated wire removed from the take-up stands are rolled onto a chock-strip in the operating aisle centrally between the two take-up stand lines. This strip serves as a temporary storage and spots the finished wire spools directly under and overhead monorail. The spools are picked up from this strip with a hair-pin hook on a trolley-mounted electric hoist and pushed to a pallet-loading area at the open end of the take-up line. Each pallet holds four spools of wire, and both spools and pallets are of such size that four spools may be placed with a single spotting of the pallet. The pallets are of special design, double-sided and made for stacking. Loaded pallets can be stacked four deep, the upper pallets resting on the heads of the spools on the lower pallets. The spool heads are made of cast steel and are of ample strength to sustain this load.

The loaded skids are picked up with an electric fork truck and moved to the loading dock at the far end of the building. Here the loaded pallet is moved directly into a motor-tractor drawn van parked at one of 2 loading doors. The van holds one day's (24 hours) output of electroformed wire.

SAFETY FEATURES

The electroforming process involves the use of chemicals all of which are dangerous when taken internally. Many are corrosive to the skin and some form poisonous vapors and gases under abnormal conditions. Specifically, extensive use of acids and cyanides demanded that extreme care be exercised in the layout of chemical storage and handling areas associated with each unit process. To aid in preventing any accidental mixing of these compounds and the possible generation of lethal hydrocyanic acid gas, a number of safety features are embodied in the design of the machine. The more important of these features are the following:

1. Individual pits for isolating the storage, mixing and filtering equipment associated with each unit process have been installed. By this action any overflow from a storage reservoir or solution leak in the handling equipment of any unit process will be confined to its own pit and may be neutralized and flushed to sewer.
2. Whenever acid and cyanide compounds occupy adjacent locations in the chemical storage and handling area, chemically resistant protective barriers have been erected between them.
3. Air wipers, water wash cells, and steam wipers have been installed in the processing sections of the machine between each acid and cyanide

bath to effectively remove from the wire any dragout from one bath before it proceeds to the next.

The safety features thus far discussed are installed with a view toward preventing the accidental generation of hydrocyanic acid gas. Additional measures have been adopted to further protect the employee. Limited admission to the electroforming building is enforced and authorized identification cards are issued only to engineering, operating and maintenance personnel assigned to the process. By such action the possibility of injury to the unacquainted observer is eliminated. Protective face shields are required equipment, they are stored on racks at the main entrance to the building and are available to all personnel. Safety showers have been strategically located throughout the building. Gas masks have been provided and are readily available at three locations. These masks are the self generation type, they will provide oxygen regardless of the nature of the surrounding atmosphere. Emergency machine shut-down buttons have been installed at four positions in the building. Finally, an evacuation alarm system has been installed for use of personnel in the building in case of an accident or failure of equipment that would result in the generation of hydrocyanic acid gas. The system includes three push-buttons under glass, one at each entrance to the building and one on the machine control consoles. A siren is provided and is audible throughout the building. Operation of any one of these buttons will trip the 480-volt substation circuit breakers for all power feeders with the exception of the lighting and ventilation circuit breakers and completely shut down both machines including all associated auxiliary machinery. The siren will be sounded and annunciator drop will be operated together with a bell and indicating light in the Firehouse and Watch Service Organization.

PERSONNEL REQUIREMENTS

It may be of interest to note that in planning this installation careful attention has been given to the increasing shortage of labor and to the provisions of automatic control and mechanical aids to reduce the labor requirements.

No standard measuring stick being available to determine the result of the efforts to reduce labor, it can be said that in a building 340 feet by 91 feet containing two million dollars worth of equipment capable of manufacturing two million dollars worth of product a year on a 350-day 3-shift basis the total operating force is 21 (five men per shift plus a material handler on the day shift).

The project is essentially chemical requiring a considerable amount of both electrical and mechanical engineering and may be regarded as a

good example of what can be done by a team of interested and competent engineers. The entire project from conception thru development of process and design and construction of equipment was conducted by Western Electric engineers. The product was specified and checked as to properties and electrical and mechanical requirements by Bell Telephone Laboratories.

The process as developed and operated is probably applicable to a number of other Bell System requirements which will be investigated and will form the basis for future expansion.

Semiconductor Diode Gates

By LUTHER W. HUSSEY

(Manuscript received May 22, 1953)

Diode gates, which are the body of modern pulse communication and computing devices, are discussed. Methods of analysis, by which practical first order design is possible, are given with experimental verification. The general properties of the gates, both virtues and defects, are noted and methods shown for minimizing the defects.

INTRODUCTION

Semiconductor diodes are old from the viewpoint of communication engineering. The crystal detector was a fundamental part of the early radio receiver. It was also a troublesome part. The crystals were highly variable and unreliable and there was little theoretical understanding of their physics which could be used as a basis for successfully exploring and controlling their characteristics. While the semiconductor continued to be useful in the field of power rectifiers and the copper oxide modulator later found extensive use in carrier telephone systems, the development of the vacuum tube into a relatively reliable device, with controlled characteristics, tended to eliminate the semiconductor from most of the communication field.

The copper oxide modulator, in a carrier telephone system, is an early example of the type of application which could bring the semiconductor back into competition with the tube rectifier. In a radio receiver, the difference in unit cost, power consumption, space requirements, and maintenance expense, between a tube detector and a crystal, may be small compared with the engineering advantages; but in a telephone plant, with its multiplicity of units, each of these small increments of cost may be integrated up to a major item, and may determine the economic feasibility of the whole system.

The need for better semiconductors than copper-oxide, when carrier frequencies moved up into the megacycle range, was a major stimulus to continue research in the semiconductor field. Within the last few years, physicists, using their recently acquired understanding of the structure

and mechanisms involved, have returned to other semiconductors — particularly to silicon and germanium to obtain great improvements.

The interest in diodes was undoubtedly a factor in the chain of investigations which culminated in the invention of the transistor and the development of the transistor has in turn accelerated the improvement of the diode. The development of the transistor also has greatly increased the demand for diodes. Replacing vacuum tubes by transistors, in modern pulse communication and computer system design can expand the potentialities of such systems since such systems, even more than the telephone carrier, use large numbers of elements and the increments of cost (initial unit cost, power consumption, space requirements, maintenance) all become vital. In a typical system there may be a dozen diodes to every transistor or vacuum tube. In a vacuum tube system the use of tube diodes is generally undesirable. In a transistor system it is absurd.

FIELD OF APPLICATION

These two new and rapidly expanding applications of diodes (pulse communication and computing) are amazingly simple in their basic ideas and circuit building blocks. Practically all they do is:

1. Generate pulses or accept them from another source and regenerate them.
2. Store pulses.
3. Route pulses to a desired output.

The complexity of the modern computer with its thousands of tubes or transistors and diodes lies in the number of operations, not in the individual operation itself. The modern computer is capable of solving complex mathematical problems involving any of the normal mathematical operations and may do it primarily with diode networks — aided by amplifiers to compensate for losses, and delay networks, or timing devices, to insure that processes take place in the proper sequence¹. The diode networks, which are the body of a modern computer or pulse communication system, are “gates” and routing circuits which, by controlling and guiding the passage of pulses, are capable of performing all the logical processes required.

TYPES OF DIODE GATES

A “gate”, for our purposes, is an electrical device with an input, an output, and one or more control inputs. The control inputs and the other

¹ Felker, Regenerative Amplifier for Digital Computer Applications, Proc. I.R.E., 40, Nov., 1952. Chen, Diode Coincidence and Mixing circuits, Proc., 38, May, 1950.

input may be indistinguishable. When a certain combination of potentials is impressed on the controls an output appears at the output terminals. In the case of a "linear" gate the input signal has simply passed from input to output, so the output signal is approximately a replica of the input signal. In the case of a "switching" gate the output is a pulse which may have no resemblance, except in duration, to the pulses and potentials which are impressed on the controls.

There is essentially only one kind of linear gate. When a signal appears at its input, the potentials then present on the controls determine whether it shall pass to the output or be blocked. Switching gates, in spite of their apparent complexity in some particular applications, are constructed of two basic types — OR circuits and AND gates. A typical OR circuit is shown in Fig. 1. When a positive pulse is impressed on either of the input terminals (terminal 1 in the illustration) it drives the corresponding diode conducting and the pulse appears at the output (3).

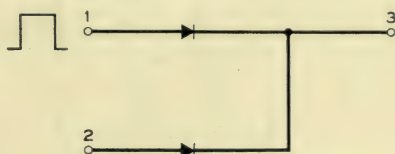


Fig. 1 — Diode OR circuit.

The signal pulse drives the other diode non-conducting, blocking that path.

An AND gate is shown in Fig. 2. The biases are so adjusted that, in the rest condition, the output voltage is zero, the diodes are conducting and the bias current from the bias current source (B) flows through the diodes and the control terminal resistors. In this case, when a positive pulse is impressed at one of the control inputs the corresponding diode is cut off and bias current ceases to flow in it. However, there is very little resulting change at the output because the bias can still flow in the second, low impedance control diode and resistance. If a second pulse is simultaneously impressed on the second control input, the second diode is also cut off and the bias current is forced to flow in the load, producing an output pulse of magnitude determined by the magnitude of the bias current and the magnitude of the load resistance.

There is a third fundamental gating concept which must also be introduced. That is the INHIBITING control. Its purpose is to prohibit an output, whatever may be done to the other controls. It is illustrated in Fig. 3. As far as control 1 is concerned, this is like the AND gate. The output, in the rest condition, is zero and the diode is conducting. There

might, in fact be two or more such controls, forming a standard AND gate. The second control shown is the INHIBIT control. It is so biased that the diode is in the high impedance, cut off condition. If let alone, it has negligible influence on the behavior of the remainder of the system. If a large negative pulse is impressed on it, overcoming the bias (C_2) the diode becomes conducting and the load remains shunted by the low control circuit impedance. Whatever the condition of the other control (or controls) very little current can get to the load and an output is inhibited.

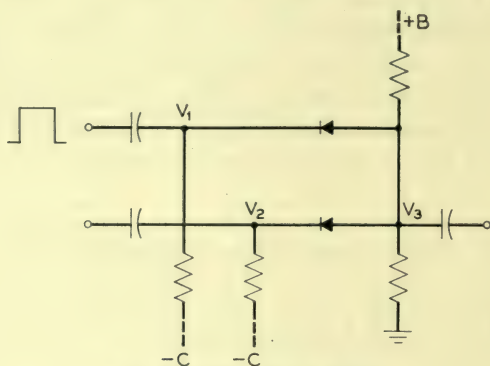


Fig. 2 — Diode AND gate.

There are several possible demands that might be put on a two control gate which can be shown in the following tabulation. In this tabulation a (1) means the presence of a pulse; a (0) means no pulse.

<i>Control 1</i>	<i>Control 2</i>	<i>Output</i>
0	0	0
0	1	0
0	1	1
1	1	1
1	1	0

These, and the corresponding relations when the two controls are interchanged, are all the available combinations of input and output. The first two are trivial — satisfied by a lack of any connection between input and output. The third demands an OR circuit. The fourth demands an AND gate. The final case requires a new configuration but can be satisfied by a combination of gates by adding the inhibiting controls. Fig. 4 shows such a gate. The boxes show gates of the type of Fig. 3, with the functional symbolism which is commonly used. A line with an

arrow aimed at the box represents a control; an arrow on a line, aimed away, represents an output connection; a control input with a small semi-circle represents an inhibiting control. This combination satisfies what is called the AND NOT requirements. A pulse appearing at input (1) will pass gate (1) if there is no pulse simultaneously on gate (2). Similarly, a pulse on input (2) will pass gate (2). If pulses appear simultaneously on both inputs, each will inhibit the passage of the other and there will be no output.

In practice there are many complications encountered in using these

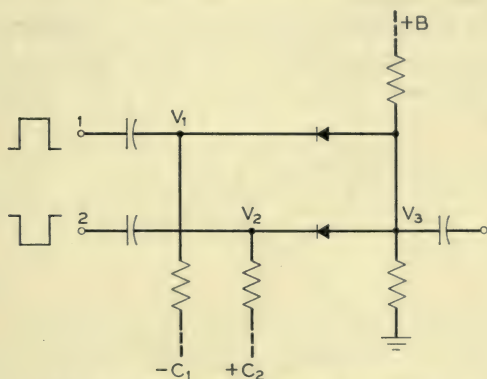


Fig. 3 — Gate with inhibiting control.

gates. In particular, networks using the INHIBITING control can, in the simple form show, get into difficulty due to the fact that pulses may not be exactly simultaneous. These difficulties are overcome in practical circuits by making part of the control voltages essentially dc potentials which set up the control conditions so that a final control pulse may produce an output, or be blocked, as the case requires. In a great many systems, this final pulse is an additional "clock pulse" from a master control or clock which synchronizes the sequence of operations of the entire computer. The details of how this is done and how the three types of gate may be combined is outside the scope of this paper. During the last few years there has been an extensive literature build up in the technical journals and a few books² published in this field.

TRANSMISSION GATE

Rather than analyze in detail the many forms a gate may take, only two forms will be discussed here. The methods (and in many cases the

² Keister, Ritchie and Washburn, *The Design of Switching Circuits*, Van Nostrand. Hartrie, *Calculating Instruments and Machines*, Univ. of Illinois Press.

results) can readily be applied to the numerous variants which may be encountered. The first is a form of "linear" or "transmission" gate. As previously stated, this gate has one or more control inputs, a signal input, and an output. When the control input enables the gate, a reasonably accurate replica of input signal should appear at the output. When the control input disables the gate, transmission of a signal should be effectively suppressed. Fig. 5 shows a form of this gate with a single control. This differs from the gates previously shown in that there is a diode in series with the output. Since this gate has superior discrimination and impedance characteristics, compared with the simpler forms it seemed desirable to analyze it as the typical transmission gate.

In this circuit, I_o , G_o represent the signal generator as a current generator with internal conductance G_o . Since the transmission proper-

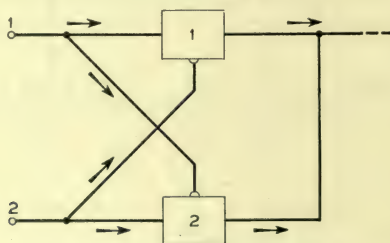


Fig. 4 — Combination of inhibited gates into AND NOT gate.

ties are of major interest, the load has been assumed to have the same conductance as the signal generator. A constant current bias I_b is impressed³ at the midpoint of the gate. The control input voltage is represented by E_b . The internal control generator conductance, which should be large, is assumed to be included in the corresponding diode conductance, for computing purposes. The diodes are assumed to have a large conductance G , or a small conductance g depending on whether they are forward biased or reverse biased. Fig. 5 shows the gate enabled, with the series diodes in the conducting state and the control diode in the reverse bias, or non-conducting condition. If the G and g are interchanged the gate is in the disabled condition.

The relative magnitudes of V_o , V_1 , V_2 evidently determine whether the diodes are biased forward or backward. Their magnitudes are immediately obtainable. The equations for the case shown in Fig. 5 are:

³ In case the internal conductance of an actual bias source is not sufficiently small to be neglected, its main effect will be on transmission loss. In computing the loss, the bias conductance may be added to the conductance of the control connection.

$$(G_0 + G)V - GV_1 = I_0 \quad (1)$$

$$-GV_0 + (g + 2G)V_1 - GV_2 = I_b + gE_b \quad (2)$$

$$-GV_1 + (G_0 + G)V_2 = 0 \quad (3)$$

Their solutions are:

$$V_0 = \frac{I_0 \left(g + G + \frac{G_0 G}{G_0 + G} \right) + (I_b + gE_b)G}{gG_0 + gG + 2G_0 G} \quad (4)$$

$$V_1 = \frac{I_0 G + (I_b + gE_b)(G_0 + G)}{gG_0 + gG + 2G_0 G} \quad (5)$$

$$V_2 = \frac{G}{G_0 + G} V_1 \quad (6)$$

The corresponding equations for the disabled gate are obtained by interchanging g and G in the above.

ENABLED GATE

Considering the enabled gate first, it should be evident from the figure that there are four requirements if the gate is to be properly enabled:

$$V_1 > V_0 \quad (7)$$

$$V_1 > V_2 \quad (8)$$

$$V_1 < E_b \quad (9)$$

$$V_1 > 0 \quad (10)$$

V_1 is a positive voltage, so equation (6) insures that (8) will always be satisfied.

Putting the values of V_1 and V_0 in (7) gives

$$(I_b + gE_b) > \left(\frac{g}{G_0} + \frac{G}{G + G_0} \right) I_0 \quad (11)$$

In the case that g is much smaller than G_0 (which is normally true) gE_b is effectively the current from a constant current control generator and the above inequality has a simple interpretation:

In determining whether (7) is satisfied, and the input diode held conducting, the above inequality compares the total current which the bias and control generators put into midpoint (V_1) of the gate with the

maximum current which the signal generator could put into the same point when that point is grounded. If the control and bias current sum is larger, the current in the input diode cannot reverse.

The inequality (10) is to insure that the output diode remains conducting. Substituting the value of V_1 in it gives:

$$I_b + GE_b > \frac{-G}{G_0 + G} I_0 \quad (12)$$

The only way that the output diode could be cut off (with positive bias and control) is by a large negative signal current. The above inequality requires:

To hold the output diode conducting, the sum of bias and control generator currents must be greater in magnitude than the maximum negative signal current that the signal generator could put into the midpoint when the midpoint is grounded.

A zero potential on the midpoint is the boundary condition between the diodes being conducting or non-conducting. The two inequalities together compare the currents that the generators can put into the grounded midpoint. They require: The sum of bias and control generator currents should exceed in magnitude the maximum current of either polarity, that the signal generator can put into the grounded midpoint.

There remains the inequality (9) which is necessary if the control diode is to remain non-conducting. This gives:

$$\frac{GI_0}{G_0 + G} + I_b < 2 \frac{G_0 G}{G_0 + G} E_b \quad (13)$$

This compares the same bias and signal generator currents with the current which would flow in the input and the output circuit if V_1 were replaced by E_b . If the inequality is satisfied V_1 can never get as large as E_b and the control input diode remains cut off.

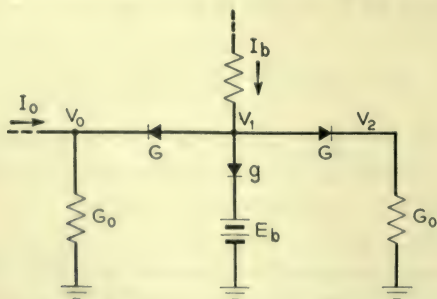


Fig. 5 — Transmission type diode gate.

DISABLED GATE

The condition under which the gate is held disabled is much simpler than the enabling conditions. All that is necessary is that V_1 be held more negative than the most negative signal generator voltage:

$$V_1 < V_0 < 0 \quad (14)$$

If equations (4), (5), and (6) are used in the disabled case (i.e., with the g 's and G 's interchanged), the condition can be obtained:

$$\frac{1}{G} I_b + E_b < \frac{1}{G_0} I_0 \left(1 + \frac{gG_0}{(G_0 + g)G} \right) \quad (15)$$

When g is much smaller than G_0 , so that the signal generator is effectively open circuited, this is a direct comparison of voltage components on the two sides of the input diode. The first term on the left, is the positive voltage at the midpoint (V_1) due to the bias current I_b . The term on the right of the inequality is approximately the open circuit signal generator voltage. The inequality says that E_b must be sufficiently negative to overcome the positive bias and hold the midpoint more negative than the most negative signal voltage.

OPTIMUM CONDITIONS

What constitutes optimum conditions depends on the particular requirements of the associated circuit. It is, however, always true that the diode conductance ratio should be as large as possible:

$$G/g \text{ as large as possible.}$$

For *minimum power loss* the gate, which is a resistive T network, to the approximation considered here, should be terminated in its characteristic conductance:

$$G_0 = \sqrt{\frac{gG}{2 + g/G}} \quad (16)$$

For *small voltage loss* the terminating conductance should be small:

$$G_0 \ll G$$

For good discrimination (large loss in the disabled state) the terminating conductance should be large:

$$G_0 \gg g$$

These all add up to the general requirement:

$$g \ll G_0 \ll G \quad (17)$$

and the particular choice of G_0 is generally a compromise between good discrimination and low transmission loss.

PEDESTAL

A major defect of diode gates is an output voltage variation produced by control voltage changes. The output voltage is very small and negative when the gate is disabled, but when the control voltage switches the gate to the enabled condition a positive voltage, called the pedestal, appears at the output. From (5) and (6), with $I_0 = 0$, it is

$$V_2 = \frac{1}{G_0} \frac{I_b + gE_b}{2 + \frac{g}{G_0} + \frac{g}{G}} \quad (18)$$

Since g is small, this approximates the voltage due to half the bias and control currents flowing in the output. The signal output is superposed on this pedestal and may swing from zero to twice V_2 . If the frequencies involved in the signal and the control voltages are widely different, this pedestal is not important since it can be filtered out from the transmitted signal but it is a serious output distortion in other cases where signal frequency components from the control pulse may be transmitted as a spurious signal.

EXPERIMENTAL CHECK

In the preceding discussion it was tacitly assumed that a diode switches between two constant conductance values. On a dc basis, the equations are still valid with a variable conductance, except for the signal loss relations. In computing the small signal loss, the conductances chosen should be the dynamic or differential conductances at the particular bias currents chosen.

The following is an example of the kind of experimental checks obtained in which, using 400B diodes, the actual behavior of a gate was tested on a dc basis. The particular units had approximately an impedance ratio (at about 5 volts) of 20,000/200. The load conductance was chosen (somewhat arbitrarily) as the geometric mean of the diode conductances, giving:

$$G = 5 \cdot 10^{-3}$$

$$g = 5 \cdot 10^{-5}$$

$$G_0 = 5 \cdot 10^{-4}$$

If the maximum signal generator current is chosen as

$$I_0 = 5 \cdot 10^{-3}$$

inequalities (11) and (13) take the form:

$$I_b + 5 \cdot 10^{-5} E_b > 5.05 \cdot 10^{-3}$$

$$-1.1 I_b + 10^{-3} E_b > 5.0 \cdot 10^{-3}$$

No negative signal voltages were used, so inequality (12) is not involved. The above inequalities limit I_b and E_b as shown in Fig. 6. I_b and E_b must be chosen from the shaded area. The values chosen were $I_b = 5 \text{ ma}$ and $E_b = 15 \text{ volts}$.

Comparing experimental results with analytic, gives:

Voltage loss 1.0 db (computed 1.6 db)

Pedestal 5.05 volts (computed 5.45 volts)

Further experimental results, which are all in reasonable agreement with expectations are given on Figs. 7, 8 and 9. Fig. 7 shows how the pedestal voltage varies with gate voltage (E_b). This is also a measure of the load capacity, since the signal output can swing from zero to twice the pedestal. The useful range is above the point where the pedestal ceases to increase with gate voltage. In this range the control diode is cut off. Figure 8 shows the pedestal against bias current I_b . The limiting is not as sharp here because the forward conductance of the input and output

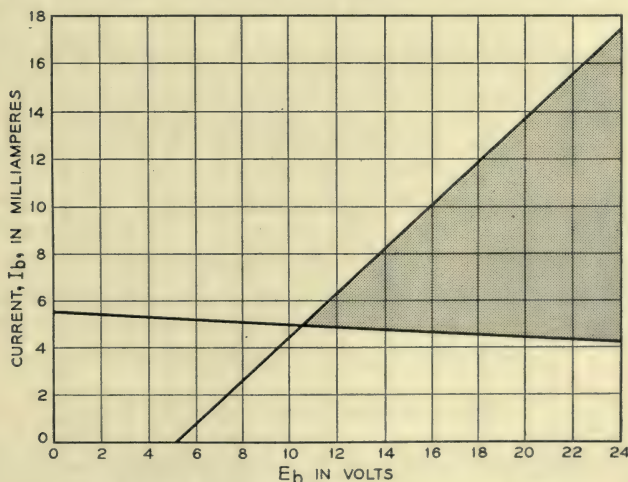


Fig. 6 — Bias restrictions on transmission gate.

diodes are involved and they do not change as rapidly with current in milliamperes as the reverse conductance does with volts. Fig. 9 shows the actual input output relation for the signal. It is linear over 80 to 90 per cent of the 5-volt range and then limits as expected.

The discrimination was not measured. It computes to better than 60-db voltage loss and discrimination of that order of magnitude has been measured in gates of this type.

SWITCHING GATE

A form of gate which is useful for pulse systems, since it lacks the pedestal, is shown on Fig. 10. This is, of course, basically the same configuration as that shown on Fig. 5, but it is operated quite differently, with pulses or dc potentials applied to the two control inputs and an output obtained by switching the bias current from flowing in a control path to flowing in the load. More specifically, if both E_1 and E_2 are sufficiently negative, diodes D_1 and D_2 are both conducting, V_b is negative, and D_3 is non-conducting. Thus practically all the bias current I_b flows in D_1 and D_2 , and V_2 is zero or slightly negative. If one of the control voltages is increased until its diode cuts off, the bias current can still flow in the other control path and the change in the output voltage is extremely small. If both the control voltages are increased until the two control diodes are cut off, then V_b becomes positive, D_3 conducts and practically the entire bias current flows in the load, producing an output voltage

$$V_L = I_b R_L.$$

The above operation gives a two control AND gate with no pedestal

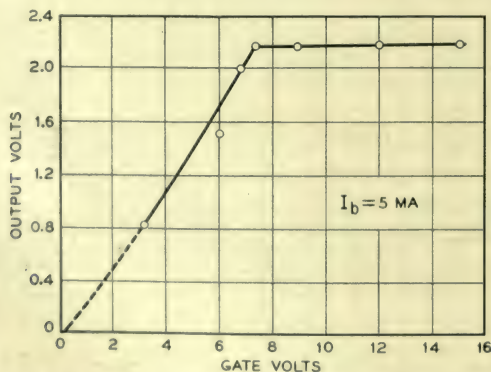


Fig. 7 — Transmission gate output (pedestal) potential versus gating control potential.

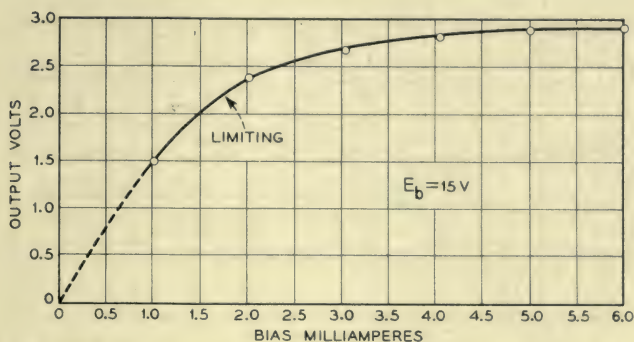


Fig. 8—Transmission gate output (pedestal) potential versus bias current magnitude.

difficulty. One of the controls could instead be used as an inhibitor. For example, if E_2 were normally biased sufficiently positive it would have no effect on the output, which would be controlled by E_1 alone. However, a negative, or INHIBITING pulse, sufficient to make D_2 conducting, would permit the bias current to flow in that path and prevent an output, whatever the state of D_1 .

This circuit could be analyzed in exactly the same manner as was done with the transmission gate. However, after a value of R_L has been chosen, a simple first approximation to a design may be carried out by assuming that the diodes are ideal, switching between zero and infinite resistance. In choosing R_L there are three major considerations:

1. R_L must be small compared with the reverse resistance of the diode, D_3 , or there may be an appreciable negative output when the gate is disabled.
2. R_L must be large compared with the forward resistance of D_3 for efficient operation of the enabled gate,—preventing an appreciable voltage loss due to the voltage drop in D_3 .
3. The peak amplitude of the output pulse is $I_b R_L$.

The value of R_L , which is chosen from the wide range of possibilities, is a matter of practical compromise, depending on the impedance levels in the system and the constant current generators which are available.

Having chosen R_L and I_b there remain only the control voltages, E_1 and E_2 . The voltages which are necessary to hold the gate enabled can be obtained by noting that (in the ideal diode case)

$$V_b = V_3 = R_L I_b \quad (19)$$

To hold the control diodes non-conducting the voltages E_1 and E_2 must be greater than V_b . This gives:

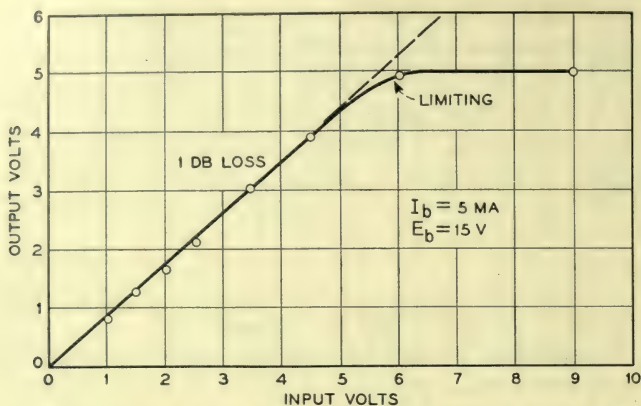


Fig. 9 — Transmission gate signal output potential versus input signal potential.

To *enable* the gate, voltages must be impressed such that

$$\begin{aligned} E_1 &> R_L I_b \\ E_2 &> R_L I_b \end{aligned} \quad (20)$$

In disabling the gate, either one of the control paths may have to carry the full bias current. If, for example, the bias current were flowing in D_2 it would bias that control path positive by an amount $R_2 I_b$. To overcome this, and keep V_b negative, the requirement is:

To *disable* the gate, voltages must be impressed such that

$$\begin{aligned} E_1 &< -R_1 I_b \\ E_2 &< -R_2 I_b \end{aligned} \quad (21)$$

These sets of conditions give the magnitudes of the biases which are necessary to hold the gate either enabled or disabled, and the difference between them is the minimum magnitude of the necessary switching pulse.

EXPERIMENTAL GATE

An example may be given, using 400B diodes. The bias was chosen as 5 *ma* and the load resistance, 2,000 ohms. The control resistances were made small, as is desirable for reasons which will be discussed later. For this gate the output voltage is 10 volts. From (20) and (21) the gate could be enabled by 10 volt positive control pulses and disabled by very small negative control values. A larger than necessary control voltage was put on control 2,

$$E_2 = 15 \text{ volts}$$

and the output measured as a function of E_1 . The results are shown on Fig. 11. Since the diodes are not ideal, there is a transition region, but, as predicted the output is very small at small negative control voltage and the output is the full 10 volts when E_1 is 10 volts.

The curve also shows what happens if one of the enabling biases are too small. A case is shown in which E_2 was only 5 volts. There is no significant difference until the output gets up to 5 volts. Above that voltage the diode, D_2 , becomes conducting and the output is clamped at that voltage.

GATE CHARACTERISTICS

The main virtues of this type of gate is that there is no pedestal and a constant amplitude pulse is produced. It is also simple and has good discrimination. There are limitations:

1. Unless a very low control path resistance is used, there is a large loss — that is the output pulse is much smaller than the control pulse. For example, if the control resistance is equal to the output resistance there is a two to one loss.

2. A rather large load is put on the control generator, partly because it must produce the enabling voltage across a small resistance and also because, in some cases the total bias current flows in the control generator output.

3. A phenomenon called "hole storage", which is present to some extent in all semiconductor diodes can make trouble. When a diode has been resting in the conducting state with a current flowing in it and the voltage is reversed, the diode does not immediately change to high impedance. A reverse current flow for a short time — up to a few microseconds. This can result in very inconvenient, spurious, output pulses being produced by a gate which is supposed to be disabled.

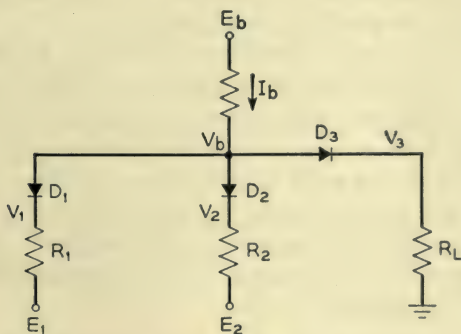


Fig. 10 — Switching type diode gate with two controls.

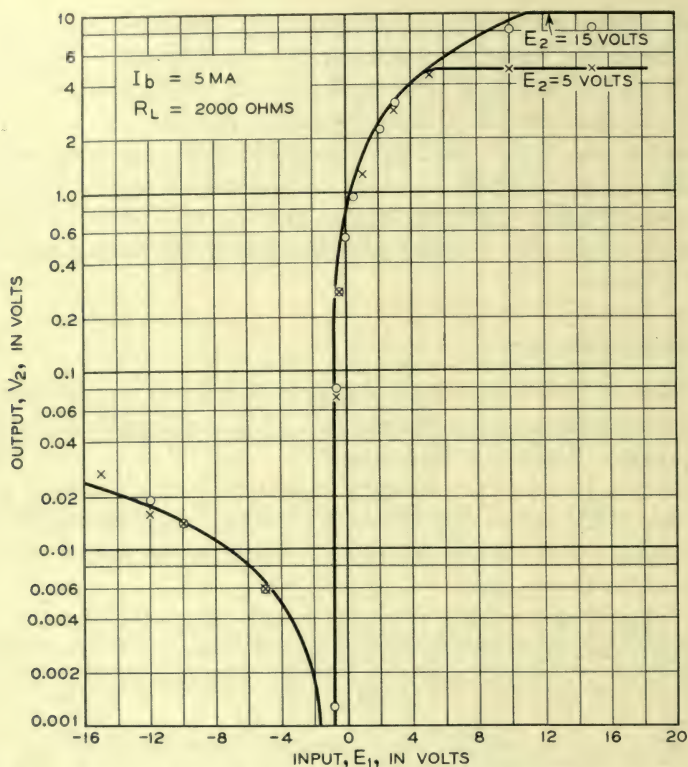


Fig. 11 — Switching type diode gate output potential versus control potential with a second control in the "enable" state.

SPECIAL CIRCUIT

It is not intended, in this paper, to attempt to list the numerous modifications which have been made of diode gates to overcome limitations and satisfy particular requirements, but it seems desirable to note an example of means to minimize the limitations.

One simple way of minimizing the difficulty, is to use point contact diodes in places where a spurious pulse could make trouble and use junction diodes elsewhere, since junction diodes have better impedance ratios but worse hole storage. For example, the output diodes in the switching gate could be a point contact unit, while the better discrimination of the junction unit made use of for control diodes.

A second means of avoiding hole storage effects is to avoid leaving diodes with large currents flowing in them, when they must be switched rapidly to the non-conducting state.

Fig. 12 illustrates both these design ideas. One of the control inputs has the control pulse impressed by means of a transformer and so has a very low dc impedance without making excessive demands on the control pulse generator. The biases are so adjusted that, in the disabled state, all the control diodes are just on the edge of conducting except the one in the transformer path. Because of the low DC impedance, practically all the bias current flows in this path. Thus there is no possibility of hole storage except in this one diode. If a positive control pulse is impressed while the potentials on the other controls are at their more negative value, the bias current just switches into those control paths; the output diode remains non-conducting and any spurious hole storage

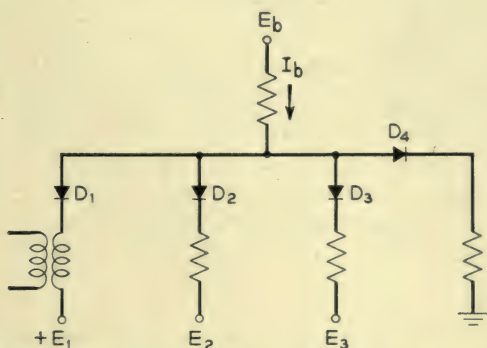


Fig. 12 — Switching type gate which minimizes "hole storage" effects.

current from the diode in the transformer control also goes into the other control paths. Since there is no storage in the other control paths, positive pulses may be simultaneously impressed on all the controls, the diodes in all but the transformer control path will immediately become high impedance and any hole storage current from the one diode will harmlessly add to the bias current flowing into the output.

Junction diodes are used in all the control inputs except the one with a transformer. A point contact unit is used here to minimize the hole storage. A point contact unit is also used in the output position. This is a critical location where good diode action is more important than a very high discrimination.

There is an additional advantage, in this configuration, that the relatively large bias current flows in the control generators only very briefly — while the disabled gate is being pulsed by the transformer control.

ACKNOWLEDGMENTS

It is impossible to give due credit to all the people who contributed to the development of the gates discussed here. Credit, however, should be given to W. D. Lewis, L. A. Meacham, and A. J. Rack who developed and explored the basic transmission gate analyzed here, and to J. R. Harris who made the modifications in the gate to avoid hole storage difficulties.

Acknowledgment is also due that part of the work reported on here was done in connection with a Joint Services contract.

A Review of New Magnetic Phenomena

By R. E. ALLEY, Jr.

Manuscript received May 25, 1953

As a result of new developments, the classical concepts of magnetic materials, characterized by hysteresis loss and eddy currents, are no longer adequate. Study of the ferrites has revealed new and important magnetic phenomena. These materials, because of their high resistivities and correspondingly low eddy currents, exhibit useful magnetic properties at frequencies well above those at which magnetic alloys are applicable. This paper reviews the new phenomena — domain wall motion and dimensional effects in the low megacycle region, and ferromagnetic resonance and the Faraday effect in the microwave region — and relates them to modern theory. Some possible microwave applications are discussed briefly.

I. INTRODUCTION

Up until a few years ago, the classical concept of magnetic materials, characterized by hysteresis and eddy current effects, was adequate for the communications engineer. Recent new and important developments, however, have made it necessary for him to have a broader knowledge of magnetic phenomena than is given by the old picture. Because of the low resistivities of existing magnetic materials, their properties at high frequencies have been dominated by eddy currents which in many cases have completely masked other magnetic effects. In contrast, the newly developed ferrites have resistivities from 10^6 to 10^{12} times greater, and eddy currents are usually negligible. The ferrites are, therefore, useful at far higher frequencies than previously available materials. Furthermore they have revealed new and important magnetic phenomena.

It is the purpose of this paper to present the modern picture of magnetism from the standpoint of the engineer. It describes the new phenomena and relates the experimentally observed behavior of magnetic materials at high frequency to the present physical theory of magnetism. The new phenomena include dimensional and domain wall motion effects in the low megacycle region and ferromagnetic resonance and the recently observed Faraday effect in the microwave region.

Although we are concerned more with magnetic phenomena than with the properties of particular materials, we will relate our discussion to ferrites, since these are the materials in which high frequency phenomena have been explored. We begin, therefore, with a brief description of these materials.

II. DESCRIPTION OF FERRITES

The term "ferrite" as used here refers to a class of ferromagnetic oxides that are structurally the same as magnetite (the naturally occurring magnetic mineral commonly known as lodestone) and as the mineral spinel from which the structure derives its name.^{1, 2, 3} These compounds form extensive solid solutions of both the substitutional and the subtractional type. Nickel zinc and manganese zinc ferrites are important examples of the substitutional type. In these, the zinc and nickel or manganese are thought to be in solid solution in magnetite (Fe_3O_4) where they have directly replaced equivalent amounts of iron in the lattice. An example of the subtractional type of solution is $\gamma\text{-Fe}_2\text{O}_3$. Here, oxygen is considered to be in solution in magnetite, not, however, having replaced iron, but having eliminated it, thus leaving vacant sites in the lattice. Magnetically, the ferrites are thought of as consisting of two interpenetrating lattices of metal ions whose magnetic moments point in opposite directions. Since, however, these moments are in general not equal, the material has a net magnetic moment.

Ferrites are manufactured by carefully mixing oxides of the constituent materials. The resulting powder is then pressed into desired shapes. These formed parts are fired at a temperature of 1000°C or more to produce the finished materials. The finished product is technically classed as a ceramic, and among its properties is extraordinarily high resistivity compared with magnetic alloys. Mechanically, ferrites have some of the characteristics of ceramics. They are extremely hard and brittle and cannot be machined by ordinary methods. They may be ground and lapped by use of abrasive cutting tools.

Experience has shown that the properties of the finished product depend upon composition (both what elements are present and in what proportions) and upon heat treating conditions (atmosphere, maximum firing temperature, and time of firing). It is apparent that, since there are so many possible variables in manufacturing procedure, one may expect a wide variety of electrical and magnetic characteristics.

Ferrites have been commercially available for several years. NiZn ferrite is being used extensively in deflection coils and flyback transformers in television sets. MnZn ferrite has found specialized but im-

portant use in inductors for networks and in transformers in telephone circuits. Magnetic recording tape makes use of $\gamma\text{-Fe}_2\text{O}_3$. An increasing amount of information of interest to the design engineer is becoming available in manufacturers' catalogs.

III. DC CHARACTERISTICS

For convenience in the following discussion, the frequency range under consideration has been divided somewhat arbitrarily according to the magnetic phenomena which have been observed. We will begin by discussing dc behavior of magnetic materials.

As far dc magnetic properties are concerned, the modern picture is the same as the classical one. The study of ferrites has revealed no essentially new phenomena, at least so far.

Hysteresis loops for typical ferrites are shown in Fig. 1, together with loops for iron and for permalloy. It will be observed that some ferrites compare favorably with permalloy with regard to hysteresis loss (proportional to the area of the hysteresis loop). Their saturation flux density is considerably lower, the maximum so far obtained being between 4,000 and 5,000.

By applying pressure to a sample and thereby introducing strain, some investigators have found it possible to produce ferrite cores having practically rectangular hysteresis loops,⁴ just as similar effects have previously been obtained with permalloy and other magnetic material.

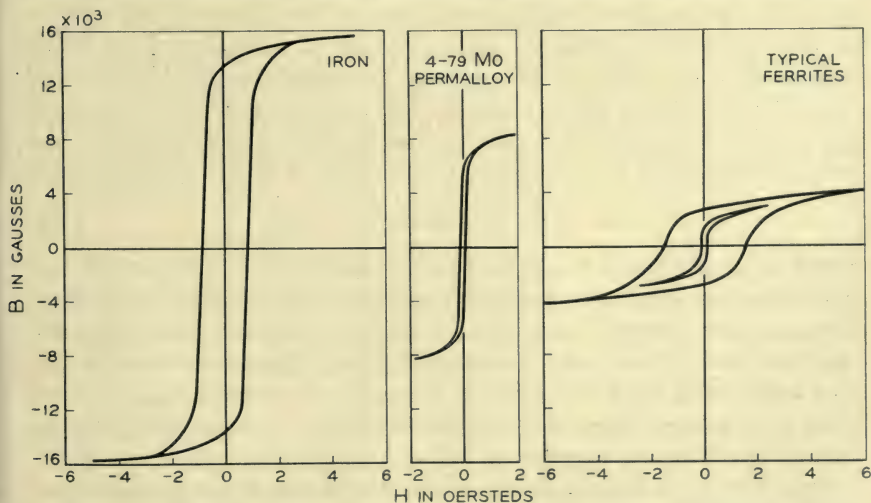


Fig. 1 — Hysteresis loops of iron, permalloy and typical ferrites. The thin ferrite loop is for a MnZn ferrite, while the larger loop represents a NiZn ferrite.

Such cores find an important application in memory circuits employed in connection with digital computers. Suitably cut ferrite single crystals also exhibit rectangular loops when properly annealed.⁵

Initial permeability of ferrites has a wide range of values depending upon the material under consideration. It may be as low as 4 for magnetite and as high as 3,000 for manganese-zinc ferrites. Curie temperature (the temperature above which the material no longer exhibits ferromagnetic properties) is around 80°C for the high permeability MnZn materials and is several hundred degrees C for the low permeability nickel ferrite. Resistivity also depends upon the composition of the material. Typical data gives values of 100 ohm-cm for a MnZn ferrite and 10^6 ohm-cm for a NiZn ferrite. The commonly used metallic magnetic materials have resistivities of the order of 10^{-5} ohm-cm. As mentioned above, this much higher resistivity is the feature of ferrites which makes possible their application at frequencies where ordinary metallic materials are generally not usable. At dc the dielectric constant of ferrites is high. Determination of dielectric constant is rather difficult, but the best measurements to date indicate values of from 10 to 30.

IV. LOW FREQUENCY PHENOMENA (0 to 1 mc)

A convenient and commonly used method of determining low frequency characteristics of magnetic materials consists in making bridge measurements of inductance (L) and effective series resistance (R) of a uniform winding placed on a toroidal core of the material. Subtraction of the dc winding resistance gives a value of resistance (R_m) which represents the core loss in the material. Permeability may be calculated from the inductance measurements. The method of analysis described by Legg⁶ may be applied to powdered or laminated alloys. The method consists essentially in determining the coefficients in the equation

$$R_m = c\mu f^2 L + a\mu B_m f L + c\mu f L, \quad (1)$$

where B_m is maximum flux density, μ is permeability, f is frequency, and e , a and c are constants. For alloys in laminated or powder form these constants are associated respectively with eddy current, hysteresis, and residual losses. From measurements of L and R_m at two or more frequencies with a fixed flux density, B_m , and two or more values of flux density at a fixed frequency, f , the coefficients e , a and c can be determined by solving the simultaneous equations obtained from equation (1).

Equation (1) is equally applicable to the ferrites at frequencies below that at which domain wall resonance and dimensional effects (discussed in Sections VI and VII) begin to appear. This frequency, which is some-

what below that indicated by f_1 in Fig. 3, depends upon the particular ferrite. The range of applicability of equation (1), therefore, varies. Regardless of frequency, the eddy current and hysteresis terms in equation (1) are valid. However, additional terms are required to cover other losses which become quite high and in comparison with which the "residual" term in equation 1 may be negligible.

Frequently, the separation of losses indicated in equation (1) is not called for. The practice then is to lump all losses together and express them in terms of the material Q , which may be a function of frequency and flux density. Here Q is the ratio of the reactance of the winding on a toroidal ring to the core loss expressed as a series resistance,

$$Q = \frac{\omega L}{R_m}.$$

The product μQ is convenient in describing magnetic characteristics. Sometimes air gaps are introduced in ferrite magnetic cores in order to provide higher coil Q 's or greater stability of ac permeability with superposed dc magnetizing force. It can be shown that the product μQ remains constant even though the core is divided by one or more air gaps. Fig. 2 shows curves of μQ versus frequency for various ferrites and also for certain other materials. An extensive discussion of methods of measurement and of results of measurements at low frequencies is given in a recent paper by Owens.⁷

It should be pointed out that eddy currents in ferrites are so small that solid shapes are usually used whenever ferrites are applicable. This is a considerable advantage, eliminating the necessity for thin tapes or insulated fine particles which are required in many applications of metallic cores. However, so far, no ferrite has been developed which has permeability nearly as great as that of some of the permalloys.

V. CHARACTERISTICS ABOVE 1 MC

So far, our discussion has covered the frequency range in which magnetic materials have traditionally found many important uses and in which the ferrites have properties generally like other magnetic materials, differing from them only in degree. We now come to the frequency range in which new magnetic phenomena have recently been observed — a range above that in which magnetic materials have heretofore been generally applicable.

In discussing the higher frequency characteristics, it is desirable to introduce a somewhat different set of parameters by which the properties

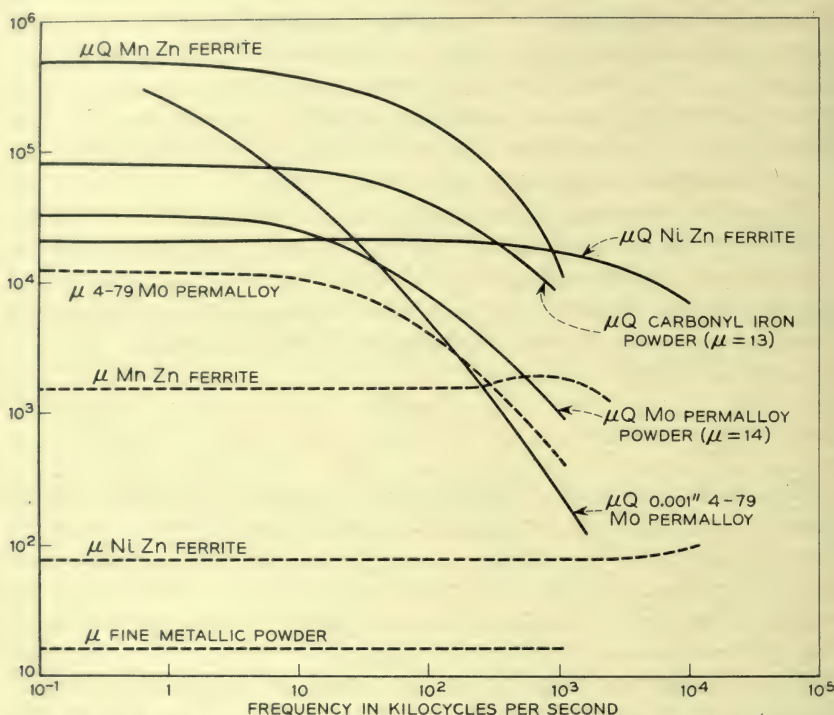


Fig. 2 — Comparison of the frequency variation of μ and μQ for ferrites and other magnetic materials.

may be described. This comes about because the ferrites, in which the new effects are observed, have both dielectric and magnetic properties. The material is most conveniently described in terms of two complex quantities: the permeability, $\mu = \mu' - j\mu''$, and the dielectric constant, $\epsilon = \epsilon' - j\epsilon''$. μ' corresponds to the usual low frequency permeability and $Q = \mu'/\mu''$. Similarly ϵ' corresponds to the usual low frequency dielectric constant, while $\epsilon''/\epsilon' = \tan \delta$, the loss tangent of the material. Thus the quantities μ'' and ϵ'' are measures of the magnetic and dielectric losses per cycle respectively, in the material.⁸ The fact that μ'' and ϵ'' represent loss *per cycle* means that much higher values of these quantities can be tolerated at low frequency than at microwave frequencies.

At frequencies above a few megacycles, it becomes very difficult to make meaningful observations on wound toroidal cores. Such difficulty may be overcome by making measurements on a toroidal sample placed in a coaxial line. Details of the experimental procedure may be found in references 9 through 12. The same procedure may be applied at microwave frequencies with waveguide used instead of coaxial line. In this case,

the sample is in the form of a slab, cut to fit snugly against the walls of the waveguide.

Fig. 3 shows in a qualitative way the behavior of a typical ferrite. In this figure we have plotted the real and imaginary parts of permeability as functions of frequency. Examination of the results obtained by various investigators for a number of different ferrites leads to the conclusion that they all behave in a fashion similar to that shown in Fig. 3. Frequency f_1 usually lies in the region between 1 and 100 mc and f_2 frequently lies in the neighborhood of 3000–4000 mc. Available data are not sufficient to provide similar curves for dielectric constant, but what there are indicate a decrease from the very high apparent value at low frequencies to a constant value of the order of 10. This decrease generally occurs somewhere around 10 mc, although it depends upon the material and also probably upon sample dimensions as will be discussed below.

The consensus is that the high experimental values of dielectric constant observed at low frequencies result from the peculiar structure of ferrites. They are considered to consist of grains of conduction material (of moderately high conductivity) separated by thin layers of dielectric material having a dielectric constant of the order of 20. Measurements on such a structure would give very high apparent dielectric constant and low Q . Such behavior was observed several years ago in samples of powdered permalloy, and has also been found in samples of plastic in which finely divided particles of copper have been dispersed.

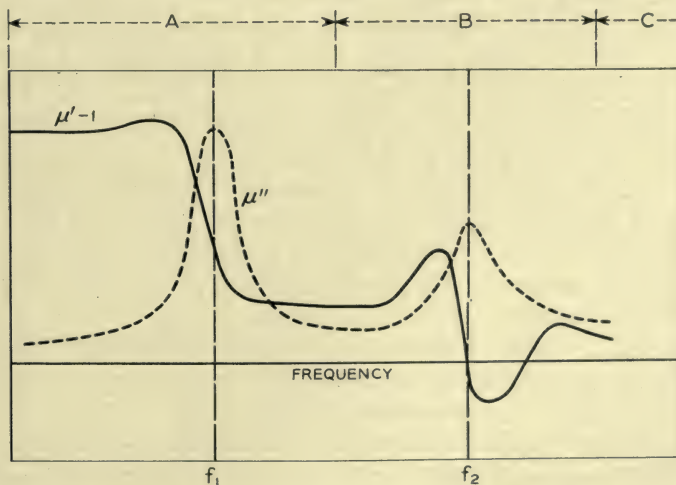


Fig. 3 — Frequency characteristics of a typical ferrite. The components of complex permeability ($\mu = \mu' - j\mu''$) as functions of frequency. The behavior in the neighborhood of f_1 is attributed to domain wall motion or to dimensional effects. f_2 represents the frequency at which ferromagnetic resonance occurs.

For convenience in the following discussion, Fig. 3 has been divided into three regions as indicated. It will be observed that there are two peaks in the curve of μ'' versus frequency, one in Region A and the other in Region B. The frequencies at which these occur are those at which relatively large amounts of energy are absorbed by the material. These two absorption peaks are due to entirely different mechanisms within the ferrite and it is, therefore, of interest to consider them separately. We begin with Region A.

The behavior indicated in Region A is typical of polycrystalline samples of ferrite (as distinguished from single crystals which will not be considered here). μ' rises somewhat above its constant low frequency value and then decreases rather suddenly. If f_1 denotes the frequency at which the peak in μ'' occurs, we find that in general a high value of μ' at low frequency is associated with a low value of f_1 , and vice versa.

At the present time, we are not sure which of two experimentally observed effects in the ferrite is responsible for the behavior shown in Region A. It is quite likely that what one observes in a given sample is actually a combination of the two effects, domain wall motion and dimensional resonance, each of which will now be described.

VI. DOMAIN WALL MOTION

The basic unit of magnetism is the spinning electron. In an atom of ferromagnetic material, there is an excess of electrons with spins in one particular direction. As a result, the atom has a net magnetic moment. Any ordinary sample of ferromagnetic material consists of many small, irregular volumes called domains, each of which may contain many atoms. Each domain is completely magnetized along some direction. Both the size of the domains and the directions of their magnetizations vary from point to point throughout a sample. In an unmagnetized material, the random orientation of individual domain magnetizations results in a mutual cancellation of their effects. However, if a magnetic field is applied, certain domains will be in a preferred orientation, having their magnetic moments more nearly in the direction of the applied field than others. These will grow at the expense of less favorably oriented domains by a process of motion of the walls separating adjacent domains. When an alternating field is applied, the walls will be subject to an alternating force which will tend to move them first in one direction and then in the opposite direction. Now it has been shown¹³ that, under these conditions, the permeability of the material is proportional to the ease of displacement of the domain walls. Therefore, if we can predict how a wall will move as the frequency of the applied field changes, we can predict how the permeability will change with frequency.

The walls have been found to exhibit properties of mass and stiffness and to be subject to damping. Thus it is possible to write an equation which describes the motion of a wall under the influence of an alternating field.^{13, 14} If x is the displacement of the wall from its equilibrium position, then

$$m \frac{d^2x}{dt^2} + \beta \frac{dx}{dt} + \alpha x = M_s H e^{j\omega t}, \quad (2)$$

where

m = mass/unit length of wall,

β = damping coefficient,

α = stiffness parameter,

M_s = saturation moment of sample,

$H e^{j\omega t}$ = applied field,

This equation should look familiar to electrical engineers. It has the same form as that for an RLC circuit subject to a sinusoidal applied voltage.¹⁵ If the damping constant is not too great, we would expect the wall motion to have a resonance at a frequency, f_0 , for which

$$\omega = \sqrt{\alpha/m}.$$

Fig. 4 shows the expected variations of μ' and μ'' in the vicinity of resonance under these conditions.

If, however, the mass of the wall is negligible, the situation is somewhat different. In this case, μ'' has a maximum at a frequency for which $\omega = \alpha/\beta$, while μ' decreased monotonically from its low frequency value, reaching $1/2$ this value at $\omega = \alpha/\beta$. It is apparent that this behavior is analogous to that of an RC circuit. It is commonly known by the term "relaxation". We see, therefore, that, depending upon wall mass, there are two possible ways in which wall motion may contribute to the behavior shown in Region A of Figure 3; namely, through domain wall resonance and through domain wall relaxation.

VII. DIMENSIONAL RESONANCE

The second phenomenon which may aid in accounting for the behavior indicated in Region A is dimensional resonance. In 1950, Brockman, Dowling, and Steneck¹⁶ reported the results of some experiments on a manganese-zinc ferrite known commercially as Ferroxcube III. Using blocks of this material, they built a closed rectangular core. Measurements on a winding placed on this core showed behavior like that of region A with f_1 lying between 1 and 2 mc. When they decreased the

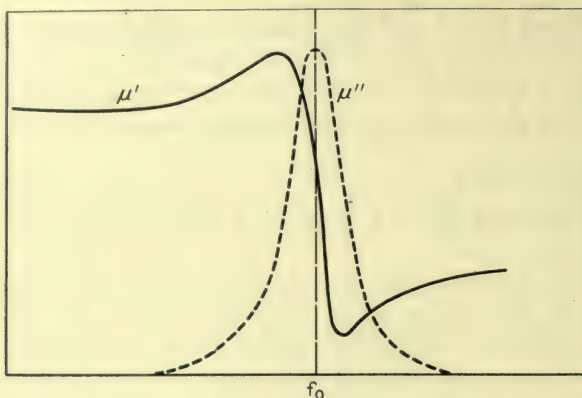


Fig. 4 — Typical behavior of the components of complex permeability ($\mu = \mu' - j\mu''$) in a ferrite in the neighborhood of domain wall resonance.

cross section area, f_1 moved to a higher value. Brockman, Dowling and Steneck attributed this to a cavity type resonance effect resulting from a high dielectric constant which gave wavelengths in the material of the same order of magnitude as the dimensions of the sample under test. Another way¹⁷ of looking at the phenomenon is to recognize that in the ferrites there are displacement eddy currents which correspond to the conduction eddy currents in magnetic alloys. Analysis based on this approach gives resonance frequencies which are the same as those calculated by Brockman, Dowling and Steneck.

In any sample it is possible that both mechanisms — domain wall motion and dimensional resonance — contribute to the behavior of the material in Region A. Furthermore, it is evident that considerable caution is required in interpreting results of experiments designed to measure μ and ϵ as functions of frequency. One must always bear in mind that what one obtains is the *effective* μ and ϵ of the sample under test, and the actual μ and ϵ for the material may be different from the observed values, depending upon the effect of sample dimensions.

It is clear that in a design problem the communications engineer must take into account the dimensions of the ferrite part as related to the permeability and dielectric constant of the material and to the frequency at which it is being used. This may impose a practical limitation on the size of a part for a particular application.

VIII. FERROMAGNETIC RESONANCE

The behavior indicated in Region B of Fig. 3 is attributed to ferromagnetic resonance. This phenomenon was first observed in magnetic metals by Griffiths¹⁸ and has been studied intensively in ferrites.²¹

Consider a single electron. Because it is a spinning charge, it has a magnetic moment which lies along its axis of spin. Because it has mass, it has mechanical angular momentum. The ratio of these two quantities is the magnetomechanical ratio, γ . If a steady magnetic field, H_0 , is applied, there will be a torque on the electron as a result of the interaction between H_0 and the magnetic moment of the electron. The electron will, therefore, precess about the direction of H_0 with a frequency which has been shown to be given by $\omega_0 = \gamma H_0$. This phenomenon is the well-known Larmor precession.¹⁹ We may say then, that the electron has a resonance frequency ω_0 .

Now suppose a sample of ferrite to be placed in a magnetic field H_0 . By virtue of the contributions of its many spinning electrons, the sample has a magnetic moment, M . If H_0 is a strong field, M will be parallel to H_0 . However, there will be, as in the case of the single electron, a frequency at which M will precess about the direction of H_0 . This precession frequency is proportional to an effective field, H_e , and to M/J , where J is the vector sum of the individual angular momenta of the electrons. H_e is a function of H_0 and also of the demagnetizing fields within the ferrite. If an alternating field of this frequency is supplied perpendicular to H_0 , then absorption will occur. The amplitude of the precessional motion will become such that the energy supplied by the alternating field is equal to the energy transformed into heat in the sample. At the resonant frequency, μ' is equal to 1 and μ'' reaches a maximum.

Although the above discussion postulates a strong external field, a more detailed analysis leads to the conclusion that resonance may be expected even with zero external field. This is attributed to the presence of internal fields which result from such things as crystal anisotropy, magnetostrictive strain, and internal demagnetizing fields in the material. These demagnetizing fields are generally the most important factor in determining the resonant frequency of a demagnetized ferrite.

A number of investigators have studied ferromagnetic resonance in ferrites. Some experimental methods and results are described in References 19 through 23. Much of the investigation has been carried out on single crystals of ferrite. In these cases, the experimental results depend upon the orientation of the crystal axis with respect to the external field. In the case of polycrystalline samples, the resonance is still present, but the resonance is not as sharp.

Most of the experiments in which ferromagnetic resonance has been studied have been with fixed frequency and varying magnetic field. Fig. 5 shows some typical results of such an experiment. From a practical experimental standpoint, this is preferable to varying frequency with a fixed magnetic field. However, it is apparent that if one holds the magnetic

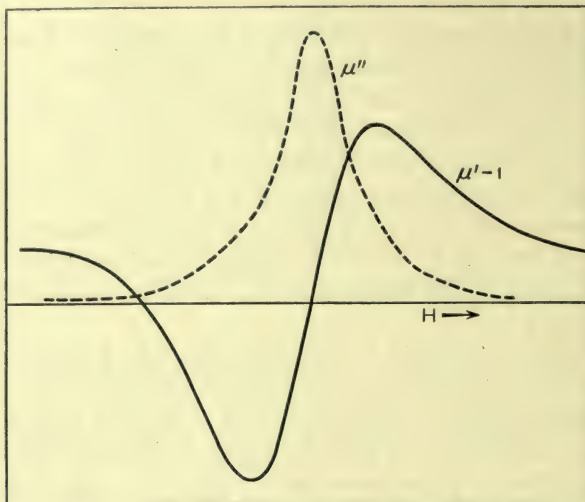


Fig. 5 — Behavior of a magnetic material in the neighborhood of ferromagnetic resonance. This represents the usual experimental situation, where frequency is kept constant and the external magnetic field is variable.

field constant at some value H_0 and varies the frequency, one will obtain curves similar to those of Fig. 5. This is indicated in Region B of Fig. 3. Furthermore, for a given sample, as H_0 increases, the frequency at which resonance occurs increases.

IX. MICROWAVE FARADAY EFFECT

If a linearly polarized wave of microwave frequency travels through a ferrite which is magnetized in the direction of propagation of the wave, the plane of polarization will be rotated. The sense of the rotation depends only upon the direction of magnetization of the ferrite and is independent of the direction of propagation of the wave. Thus the effect is anti-reciprocal.

This phenomenon, which derives its name from the analogous optical effect was demonstrated experimentally by Roberts²⁴ and has been extensively investigated by Hogan.²⁵ It occurs at frequencies above the ferromagnetic resonance frequency, that is, in Region C of Fig. 3. In principle, the effect might be expected in any ferromagnetic material but, so far, only the ferrites are sufficiently transparent to microwaves to allow the effect to be detected. The effect is illustrated in Fig. 6. The linearly polarized microwave in waveguide A passes through a transition section into the circular guide B. A tapered cylinder of ferrite is inserted in B. A solenoid, external to B, supplies a steady field parallel to the axis

of propagation. Upon emerging from the sample, the wave passes into C, a circular to rectangular waveguide transition which may be rotated for maximum transmission of energy down section C. The angle of displacement of C with respect to A is a measure of the rotation of the plane of polarization of the wave in its passage through the ferrite. The rotation per centimeter of material depends upon the longitudinal field in the sample, increasing with this field and reaching a constant value when the ferrite is saturated.

Hogan²⁵ has given a discussion of the theory of this ferromagnetic effect. The incident linearly polarized wave may be described as a combination of two oppositely rotating circularly polarized waves. The real part, μ' , of the permeability of the ferrite varies with magnetic field in a different way for the two circular polarizations, as shown in Fig. 7. The velocity of propagation of the two polarizations is therefore different, and in passing through the ferrite they will fall out of phase by an amount proportional to sample length. Upon emerging they will combine to form a linearly polarized wave whose plane is rotated with respect to the incident wave. Reference to Fig. 7 shows that the most useful region for obtaining this effect lies below the field required to produce ferromagnetic resonance (i.e., at frequencies above the ferromagnetic resonance frequency) in the region where the two curves are practically parallel. In this region the device is relatively insensitive to small field changes and is somewhat "broadband" with respect to frequency.

For the practical application of the Faraday rotation, it is desirable that the attenuation per degree of rotation be small. Attenuation varies widely for different kinds of ferrites and is a function of frequency. For

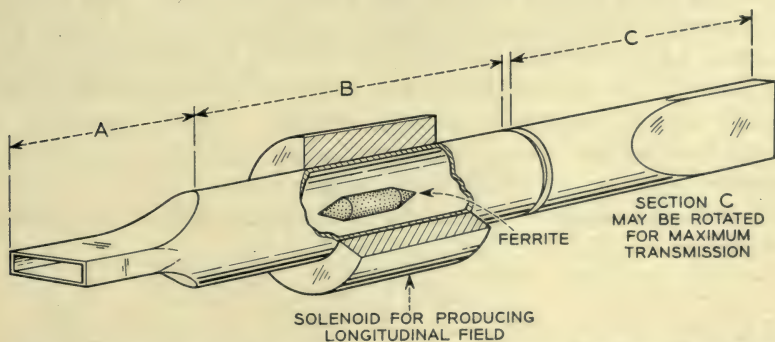


Fig. 6 — Apparatus for demonstrating the microwave Faraday effect. Energy is supplied to section A. Rotation of plane of polarization occurs in section B and is controlled by controlling the longitudinal field. Section C is rotated for maximum transmission. The angular displacement of C with respect to A is a measure of the rotation.

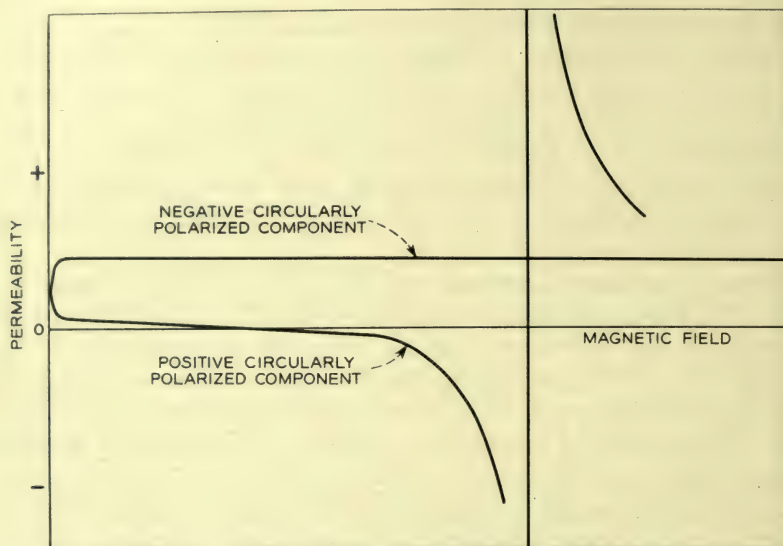


Fig. 7 — Real part (μ') of permeability of a ferrite versus applied magnetic field for the two circularly polarized components into which an incident plane wave may be resolved. The useful region for the Faraday effect is that for low H , where the two curves are approximately parallel.

one sample at 9,000 mc, Hogan found a rotation of 60 degrees per cm of path and an attenuation of about 0.5 db per cm.

X. EFFECT OF CROSS FIELD

When a steady magnetic field is applied to a ferrite in a direction perpendicular to the path of transmission of electromagnetic waves through the material, the effective ac permeability of the material varies with the applied magnetic field. For a given frequency, the permeability starts out positive. As the magnetic field increases the permeability goes through zero and approaches a large negative value as ferromagnetic resonance is reached. Above resonance the permeability is positive and gradually decreases with increase in magnetic field.

Since the characteristic impedance of the ferrite relative to an empty waveguide is

$$Z = \sqrt{\frac{\mu}{\epsilon}},$$

it is apparent that Z may be varied by changing the applied field. When μ is zero, the ferrite appears to be a perfect reflector, while when $\mu = \epsilon$, it provides a perfect match to the empty guide. Since it is possible to

vary the characteristic impedance over a wide range, this effect has possibilities of application in attenuators, modulators, and phase shifting devices.

XI. NEW MAGNETIC APPLICATIONS

The engineer is naturally interested in some of the uses to which the new magnetic phenomena may be put. Several applications will be described briefly.

1. The gyrator. This is a four pole element for which there is a 180° phase difference between the two directions of propagation. In other words, the transfer impedances in the two directions are equal in magnitude but opposite in sign. Thus the device violates the reciprocity theorem.

Hogan²⁵ built the first microwave gyrator using the arrangement shown in Fig. 8. In this device, a wave traveling from left to right has its polarization rotated 90° counter-clockwise in the twisted section and another 90° in the same direction by the ferrite — a total rotation of 180° . For a wave traveling from right to left the two rotations, that in the ferrite and that in the twisted section, cancel each other. Thus, if *A* and *B* represent points of the same phase for a left-to-right wave, they represent points of 180° phase difference for a right-to-left wave.

2. One way transmission system. If the input and output waveguides in Fig. 6 are oriented with their planes at 45° to each other and if the solenoid current is adjusted for 45° rotation in the ferrite, the result is a one-way transmission system.²⁵ Such a device is broadband. An arrangement of this sort may be employed in a microwave system to isolate the transmitter or receiver from the waveguide. It has the advantage that loss in the forward direction can be made quite small by proper choice of material.

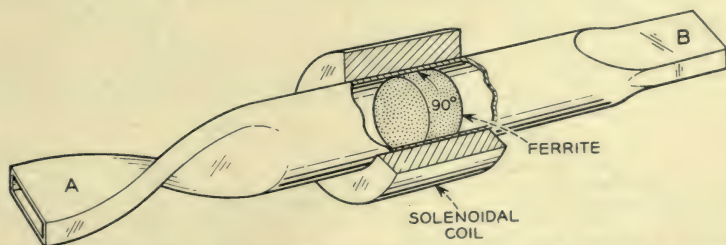


Fig. 8 — Schematic diagram of a microwave gyrator. From left to right, the plane of polarization is rotated 180° ; corresponding to a phase shift of π . From right to left, the plane of polarization is not rotated and the phase shift is therefore zero.

3. The polarization circulator. This is a modification of the one way transmission system in which there are two connections, with polarizations at 90° to each other, on either side of the ferrite rotating element. This is shown schematically in Fig. 9, along with a symbol which has been suggested for this element.²⁵ Energy sent into the device with polarization A emerges with polarization B, polarization B is rotated into C, polarization C is rotated into D, and polarization D emerges as *minus* A. One practical application of this device is as a TR box in a radar system. Another, recently suggested by A. G. Fox, is as a device for separating the various channels in a multichannel communication system. Referring to Fig. 10, the signal comes in at A. Branch B is terminated in a filter which accepts one channel but reflects the remainder of the signal which is passed on to C. Here another filter accepts the second channel but passes the remainder on to D. D in turn feeds a second circulator. This process can go on until all channels are taken care of.

4. Measurement of magnetic field strength. The phenomenon of ferromagnetic resonance suggests a means of making measurement of magnetic field strength by observing the resonance frequency for a ferrite when subjected to the unknown field.²⁶ Allen²⁷ has recently described a magnetometer in which an unknown field is measured by observing the Faraday rotation which it produces in a standard sample.

5. Other applications. There are several ways in which the interaction of the steady field with the microwave field may be utilized in designing switches, attenuators, and modulators. For example, one might set the two rectangular guides in Fig. 8 with their transmission planes at 90° . Then by varying the current in the solenoid and thereby varying the magnetic field applied to the ferrite, one may vary the amount of energy accepted by the second waveguide. This is then an electrically controlled attenuator. This same device offers the possibility of providing modulation of the microwave signal by modulating the current in the solenoid.

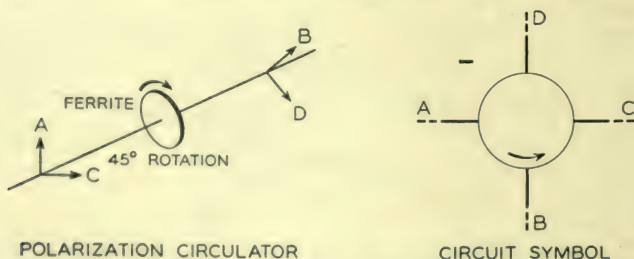


Fig. 9 — Schematic representation of polarization circulator. The ferrite is adjusted to 45° rotation by an external field, not shown. The circuit symbol for the circulator is shown at the right.

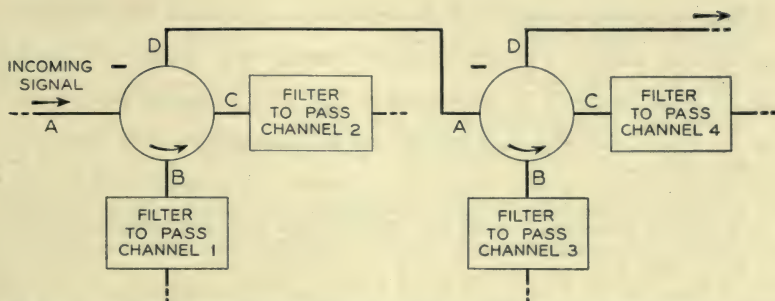


Fig. 10 — Schematic diagram showing the proposed use of circulators to separate the various channels in a multichannel communication system.

Reggia and Beatty²⁸ have recently described a coaxial line variable attenuator in which the transmission loss is controlled by variation in an external cross field.

XII. CONCLUSION

From the discussion which has gone before, it should be apparent to the communications engineer that a whole new field of applications of magnetic materials has opened up. It is therefore essential that the engineer be acquainted with the modern picture of magnetism including the phenomena which have been described here — low frequency resonance, ferromagnetic resonance, and microwave Faraday effect. Some applications have already been made of the high frequency characteristics, particularly of the Faraday rotation. Knowledge of the general high frequency characteristics of magnetic materials will enable the engineer to interpret new experimental information as it becomes available and intelligently to utilize the new materials in a variety of engineering applications.

I wish to express my appreciation to J. K. Galt, A. G. Ganz, C. L. Hogan, and V. E. Legg for several discussions which aided in the clarification of certain points described herein and to Messrs. Ganz and Legg for their careful criticism of the overall presentation.

REFERENCES

1. J. L. Snoek, Non-Metallic Magnetic Materials for High Frequencies, Philips Technical Review, **8**, p. 353, 1946.
2. A. Fairweather, F. F. Roberts and A. J. Welch, Ferrites, Reports on Progress in Physics, **15**, p. 142, London, The Physical Society, 1952.
3. R. L. Harvey, I. J. Hegyi, H. W. Levernz, Ferromagnetic Spinel for Radio Frequencies, R.C.A. Review, **11**, p. 321, 1950.
4. H. J. Williams, R. C. Sherwood, M. Goertz and F. J. Schnettler, Stressed

- Ferrites with Rectangular Hysteresis Loops, Scheduled for publication in A.I.E.E. Transactions.
5. J. K. Galt, Motion of a Ferromagnetic Domain Wall in Fe_3O_4 , *Physical Review*, **85**, p. 664, 1952.
 6. V. E. Legg, Magnetic Measurements at Low Flux Densities using the A. C. Bridge, *Bell System Tech. J.*, **15**, p. 39, 1936.
 7. C. D. Owens, Analysis of Measurements on Magnetic Ferrites, *Proc. I.R.E.*, **41**, p. 359, 1953.
 8. C. Kittel, Theory of the Dispersion of Magnetic Permeability in Ferromagnetic Materials at Microwave Frequencies, *Phys. Rev.*, **70**, p. 281, 1946.
 9. J. B. Birks, Measurement of the Permeability of Low Conductivity Ferromagnetic Materials at Centimeter Wavelengths, *Proc. Phys. Soc. (London)*, **60**, p. 282, 1948.
 10. G. T. Rado, R. W. Wright and W. H. Emerson, Ferromagnetism at Very High Frequencies. III. Two Mechanisms of Dispersion in a Ferrite, *Phys. Rev.*, **80**, p. 273, 1950.
 11. *Techniques of Microwave Measurements*, (M.I.T. Radiation Laboratory Series, **11**), McGraw-Hill Book Co., N. Y., 1947.
 12. W. H. Surber, Jr. and G. E. Crouch, Jr., Dielectric Measurement Method for Solids at Microwave Frequencies, *J. Appl. Phys.*, **10**, p. 1130, 1948.
 13. J. K. Galt, Initial Permeability and Related Losses in Ferrites, *Ceramic Age*, p. 29, Aug., 1952.
 14. R. Becker and W. Doering, *Ferromagnetismus*, Julius Springer, Berlin, 1939.
 15. M.I.T. Staff, *Electric Circuits*, John Wiley & Sons, N. Y., p. 323, 1943.
 16. F. G. Brockman, P. H. Dowling and W. G. Steneck, Dimensional Effects Resulting from a High Dielectric Constant Found in a Ferromagnetic Ferrite, *Phys. Rev.*, **77**, p. 85, 1950.
 17. P. M. Prache and H. Billottet, "Magnetodynamique des Semi-Conducteurs", *Cables et Transmission*, **6A**, p. 317, 1952.
 18. Griffiths, *Nature*, **158**, p. 670, 1946.
 19. C. Kittel, Ferromagnetic Resonance, *Le Journal de Physique et le Radium*, **12**, p. 291, 1951.
 20. H. G. Beljers and J. L. Snoek, Gyromagnetic Phenomena Occuring with Ferrites, *Philips Tech. Rev.*, **11**, p. 313, 1950.
 21. W. A. Yager, J. K. Galt, F. R. Merritt and E. A. Wood, Ferromagnetic Resonance in Nickel Ferrite, *Phys. Rev.*, **80**, p. 744, 1950.
 22. R. M. Bozorth, *Ferromagnetism*, New York, D. Van Nostrans Co., p. 803, 1951.
 23. H. G. Beljers, Measurements on Gyromagnetic Resonance of a Ferrite Using Cavity Resonator, *Physica*, **14**, p. 629, 1949.
 24. F. F. Roberts, A Note on the Ferromagnetic Faraday Effect at Centimeter Wavelengths, *Le Journal de Physique et le Radium*, **12**, p. 156, 1951.
 25. C. L. Hogan, The Ferromagnetic Faraday Effect at Microwave Frequencies and its Applications, *Bell System Tech. J.*, **31**, p. 1, 1952.
 26. Theodore Kahan, Ultra-High Frequency Magnetometer, U. S. Patent 2597149.
 27. P. J. Allen, A Microwave Magnetometer, *Proc. I.R.E.*, **41**, p. 100, 1953.
 28. F. Reggia and R. W. Beatty, Characteristics of Magnetic Attenuators at UHF, *Proc. I.R.E.*, **41**, p. 93, 1953.

The Correlatograph

A Machine for Continuous Display of Short Term Correlation

By W. R. BENNETT

(Manuscript received April 24, 1953)

An analog device has been constructed which displays short term correlation as a three-dimensional plot in which the rectangular coordinates are running time and lag time and the intensity of the pattern represents the correlation function. Preliminary tests on the properties of such a device are reported.

The place of the electrical spectrograph¹ as a signal analyzer has become well established in laboratory technology. It has occurred to many investigators however that the spectrum is not the only property of a signal which may be worthy of study, and in recent years there has been a considerable interest in other features, notably the correlation functions. On the basis of the accepted mathematical definitions the auto correlation function is the Fourier cosine transform of the power spectrum and in this sense would contain equivalent information presented in a different form. However, the mathematical definitions apply to a very long time interval and in practice we often deal with short segments of non-stationary processes. The spectrograph does not evaluate the true spectrum in such cases but gives instead a spectrum-like function of frequency which changes with the observation time. We may regard the resolving filter as performing a weighted analysis in which the most recent parts of the signal contribute most heavily to the instantaneous response. The resulting "short term" spectrum depends on the characteristics of the resolving filter as well as the signal, but over a useful range of filtering selectivity the individual peculiarities of the signal are distinguishable even though the structural background may be characteristic of the filter.

"Short term" correlation functions, in which the averaging interval is finite, have also been investigated and show phenomena analogous to

the short term spectrum. Comparison of the two short term functions is a much more involved matter than when the time interval is large in both cases. The subject has been extensively treated in the literature,² but the conclusions are still somewhat obscure. In an investigation which was started by the present author and the late Liss C. Peterson in 1950, we noted that a complete correlation analogue of the audio spectrograph had apparently never been constructed even though suggestions had appeared concerning the possibility.³ It seemed that some of the questions concerning the merits of the correlation method of analysis could not properly be answered without actually building and testing such a machine and it also appeared possible that we would thereby add another useful tool for the problems associated with speech and other signals of interest. We accordingly undertook the design and construction of a "correlatograph" following lines closely parallel to spectrographic experience in order to economize in new shop designs.

In a spectrograph, such as used in visible speech for example, a three-dimensional display is obtained in which time and frequency are two rectangular coordinates and the short term power spectrum is represented by the intensity of light or density of marking. The functional mechanism is illustrated in Fig. 1. The incoming signal is delivered to a bank of band pass filters with midband frequencies uniformly spaced throughout the frequency range of interest. The envelopes of the filter output waves are picked up in turn by a rotating switch arm to control the voltage impressed on a marking stylus. The stylus moves across the paper in synchronism with the collector arm and the paper advances after each stroke.

The analogous diagram for a correlatograph is shown in Fig. 2. We recall that the correlation of two functions is defined as the average of

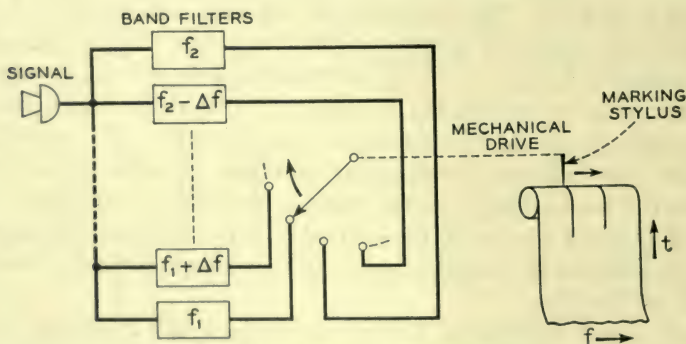


Fig. 1 — Mechanism of spectrograph.

their product with fixed time lag; i.e., when T is large, the expression

$$\psi_{12}(\tau) = \frac{1}{T} \int_0^T f_1(t)f_2(t - \tau) dt$$

gives the cross-correlation of $f_1(t)$ and $f_2(t)$, and the expression

$$\psi_{11}(\tau) = \frac{1}{T} \int_0^T f_1(t)f_1(t - \tau) dt$$

gives the autocorrelation of $f_1(t)$. In short term correlation T is finite. In Fig. 2, the different lag times are obtained by a tapped delay line and each tap is followed by its own multiplier and integrator. The integrated values are picked up by a rotating switch arm to control the marking stylus voltage as in Fig. 1. The rectangular coordinates are now t and τ instead of t and f . The marking intensity represents the correlation function.

In actual spectrographs it is usually found expedient to replace the bank of filters by a single filter and use a swept frequency oscillator and modulator to heterodyne the signal across the filter band. A rotating condenser plate tuning the oscillator thus takes the place of the rotating switch. The sweeping frequency must not change too rapidly for the analyzing filter to respond adequately, and also must not change so slowly that short signal bursts are imperfectly registered. A preliminary recording of the signal wave with a subsequent reproduction at a different

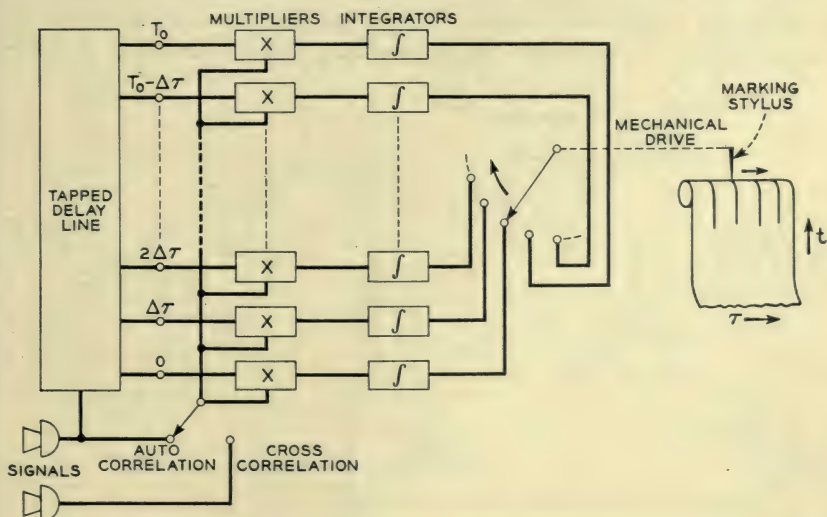


Fig. 2 — Mechanism of correlatograph.

speed forms a practical technique for securing the desired resolution in both frequency and time. Likewise in the correlatograph we can dispense with individual multipliers and integrators if we are willing to give up some of the otherwise available integration time for each τ value. One multiplier and one integrator are then put in series with the switch arm and the contacts go directly to the delay line taps as shown in Fig. 3. A complete analogy with the simplified spectrograph would require a line of variable delay with a single output tap. This could be done with magnetic recording and moving heads, thereby eliminating the rotating switch. We felt however, that the fixed delay line would furnish the more stable and accurate component and chose the arrangement of Fig. 3. An auxiliary recording and reproducing process made it possible to accommodate a wide variety of signals with the one delay line.

DETAILS OF APPARATUS

Fig. 4 shows the arrangement of apparatus chosen. Our program was aimed at signals such as might occur in a nominal speech band extending from 200 to 4,000 cps recorded on magnetic tape at 15 inches per second. The reproducing element was a spinning double-ended pickup coil which successively scanned a one-inch loop of tape with one end beginning its scan just as the other end left the tape. The coil made 60 revolutions per second and hence the reproducing speed was 120 inches per second or eight times the recording rate. Our speech band was thus made to occupy the range from 1,600 to 32,000 cps, and this was therefore the range chosen for the delay line. A recording speed other than 15 inches

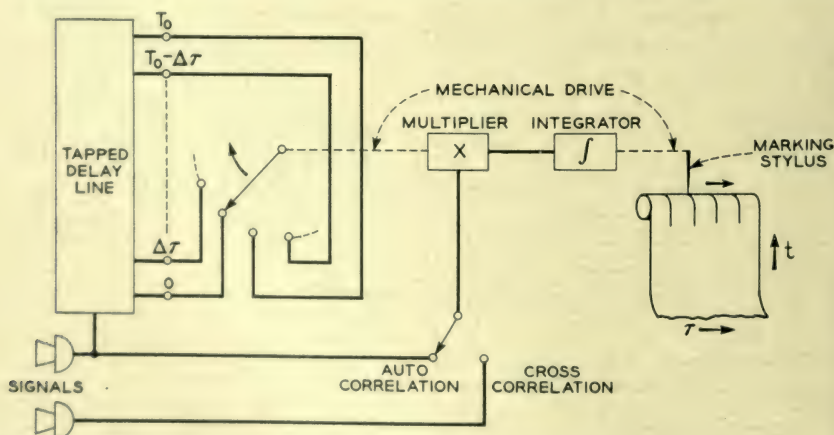


Fig. 3 — Mechanism of correlatograph with common multiplier and integrator.

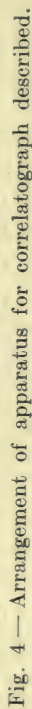


Fig. 4 — Arrangement of apparatus for correlatograph described.

per second brings a correspondingly different signal band into the range of the delay line. The magnetic tape advances relatively slowly during the scanning process and for any one scan of a one-inch section may be regarded as stationary. The revolving pick up thus delivers segments of signal $1/120$ second (8.33 ms) long for the correlation analysis. Our delay line consists of 1.5 ms total delay with 200 taps spaced $7.5 \mu\text{s}$ apart. During the first 1.5 ms of each scan, the delay line contains parts of the response from two successive scans and hence is not suitable for correlation measurement for lag times corresponding to all taps. We make provision therefore for excluding this interval from the analysis and use only the last 6.83 ms of each scan. The rotating switch advances one tap on the delay line for each one inch scan, so that the value of short term correlation corresponding to one value of lag time is computed every 8.33 ms. 200 values are computed in 1.67 sec after which a time of 0.33 sec is allowed for the return of the marking stylus to its initial position. The rotating switch thus makes one revolution in two seconds but the last sixty degrees of the revolution are not used for display of correlation. The recording paper advances 0.01 inch after each stroke of the stylus. The rate of advance of the signal tape is adjustable by means of a gear train. In terms of the original signal wave, one second of recorded time is represented by the distance the paper moves in one second (0.005 inch) multiplied by the ratio of recording speed to speed of tape advance past the scanning head.

The tape scanning mechanism with the synchronized motion of stylus and paper is due entirely to I. E. Cole, who designed this part of the system and supervised the necessary shop work. Mr. Cole also cooperated in the choice of a design plan for the rotating switch, which was manufactured to meet our special requirements by Applied Science Corporation of Princeton, N. J. The switch output is followed by an electronic gating circuit which removes the effect of time jitter in the beginnings and ends of the contact intervals and trims off the previously mentioned 1.5 ms interval during which the tail end of one scan of the tape loop remains in the delay line. The gating wave is generated from the common 60-cycle power supply which drives all the mechanical apparatus. The output of the electronic gating circuit consists of segments of signal 6.83 ms long with 1.5 ms separation and with the delay increasing in steps of 7.5 microseconds between one segment and the next. This constitutes one input to the multiplier; the other input is the undelayed signal in the case of autocorrelation or an independent signal for cross-correlation.

The multiplier consists of a bridge of germanium varistors with the two

inputs applied across the two diagonals and the output taken off one pair of input terminals through a low pass filter. Each varistor is operated in a substantially square law range. If A and B represent the two inputs, we try in effect to produce an output proportional to

$$(A + B)^2 - (A - B)^2 = 4AB.$$

The necessary conditions are conveniently expressed in terms of sine wave inputs in order that analyzer measurements may be used as a check on accuracy. Let $P \cos pt$ represent a typical component of the signal impressed as one input to the multiplier and $Q \cos qt$ a typical component simultaneously applied to the other input terminal. The product is

$$(P \cos pt)(Q \cos qt) = \frac{PQ}{2} \cos (p + q)t + \frac{PQ}{2} \cos (p - q)t.$$

Since our purpose is to integrate the output of the multiplier over a time interval relatively long compared to the periods of components within the signal band, we have no interest in the product component of frequency $p + q$ and in fact filter out such components immediately along with the original signal components to prevent loading the product amplifier with unessential waves. A significant test on the accuracy of the multiplier is therefore the fidelity with which the amplitude of the difference frequency term $\cos (p - q)t$ follows the product of the amplitudes of $\cos pt$ and $\cos qt$. This is not sufficient in itself however because it does not give a check on the balancing out of the squares of the individual inputs. To test the latter we superimpose the two components $P \cos pt$ and $Q \cos qt$ on one input circuit with no signal applied to the other input and measure the output component of frequency $p - q$. We also repeat the measurement with the two sine waves impressed on the second input and nothing on the first input. Typical results are shown in Fig. 5. The varistors were selected by R. R. Blair from persistent screen cathode ray tube displays of the characteristics. The square law region is enlarged and the output increased by applying a direct current bias to the bridge through a series resistance. This form of multiplier copies a design worked out by R. R. Riesz for a different purpose.

The product output of the multiplier is weak at best because a relatively small range of varistor inputs fit the necessary law. A fairly high gain product amplifier is therefore provided. Fortunately we do not have to amplify a band which extends all the way down to zero frequency. The significant component when calculating the autocorrelation function of $P \cos pt$ for lag time τ is $(P^2/2) \cos p\tau$, which is constant only when $p = 0$ or τ is constant. Our lowest value of p corresponds to 1600 cps. The value

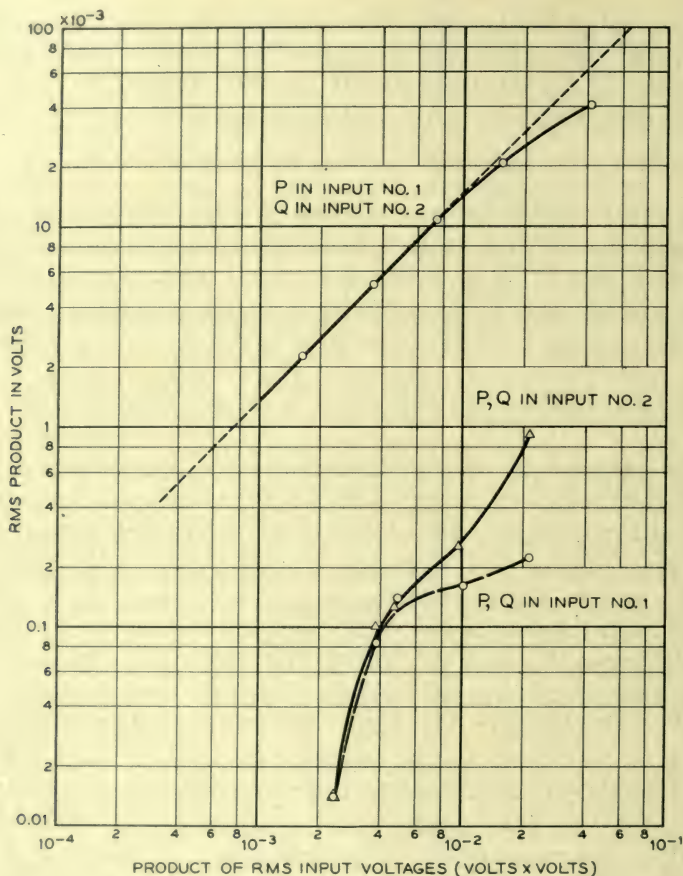


Fig. 5 — Performance curves of multiplier.

of τ increases in small steps from 0 to 1.5 ms in $5/3$ sec. and therefore can be approximated by $\tau = bt$, where $b = .0015/(5/3) = 0.0009$. Hence the lowest product frequency is roughly $1600 \times 0.0009 = 1.44$ cps. The actual low frequency cutoff of the amplifier is made about 0.1 cps to allow for changes in the signal components and to preserve good transmission within the nominal band. The upper cutoff frequency is likewise made somewhat higher than the nominal value of $32,000 \times 0.0009 = 28.8$ cps.

The integration is performed by a series condenser and shunt resistance at the amplifier output. A good approximation to integration is accomplished by a large time constant such that the indicial admittance

remains linear during the integrating interval. Such a long time constant would carry over too much charge from previous intervals if continuous integration were permitted so a shorting relay is provided to discharge the condenser quickly to ground after the integrated value is sampled. The sampling is done by a two-way clamp circuit with the timing pulse generated from the trailing edge of the pulses which gate the switch outputs. The shorting relay operates directly from the 60-cycle supply and is of a type specifically designed to give a brief closure of about one ms. every half period of the driving wave. The sampled outputs of the integrator are applied to a balanced modulator to which is also applied a 6-kc carrier. The resulting double sideband suppressed carrier wave is amplified to form the marking voltage applied to the stylus. Instantaneous compression is incorporated in the marking amplifier to extend the range of input magnitudes which are encompassed by the relatively narrow recording range of the paper. The stylus responds equally to positive and negative voltages and does not resolve the individual high frequency oscillations. The result is like full wave rectification of the correlation functions.

Grateful acknowledgement is given to A. J. Rack, A. E. Johanson and P. A. Reiling for suggested physical configurations and design information suitable for the circuits which generate the various control pulses, and which sample and hold the integrated outputs. G. W. Blake tested and adjusted these circuits after they had been constructed by the wiring shop. Performance runs were made by F. H. Tendick and N. K. Poole. Also at various stages of the project assistance was given by W. A. Klute, F. W. Kammerer and R. L. Carbrey. We have also received helpful advice from many other associates too numerous for explicit mention.

The 200-tap delay line was constructed by the transmission networks department. It consisted of one low pass filter section per tap with mutual inductance between sections to maximize the linearity of the phase curve. The taps were taken off high impedance shunts with the output fraction tapered down the line to give constant loss along the taps. C. E. Jakielski planned, assembled, tested, and adjusted the delay line. The intricate task of connecting the 200 output taps to the corresponding 200 contacts of the switch was performed by M. Biazzo. The delay line and switch appear in the photograph, Fig. 6. Additional electronic apparatus not shown in this photograph was placed on independent panels for experimental convenience but can be arranged in a chassis in the same cabinet with the other apparatus, now that the components have been determined.

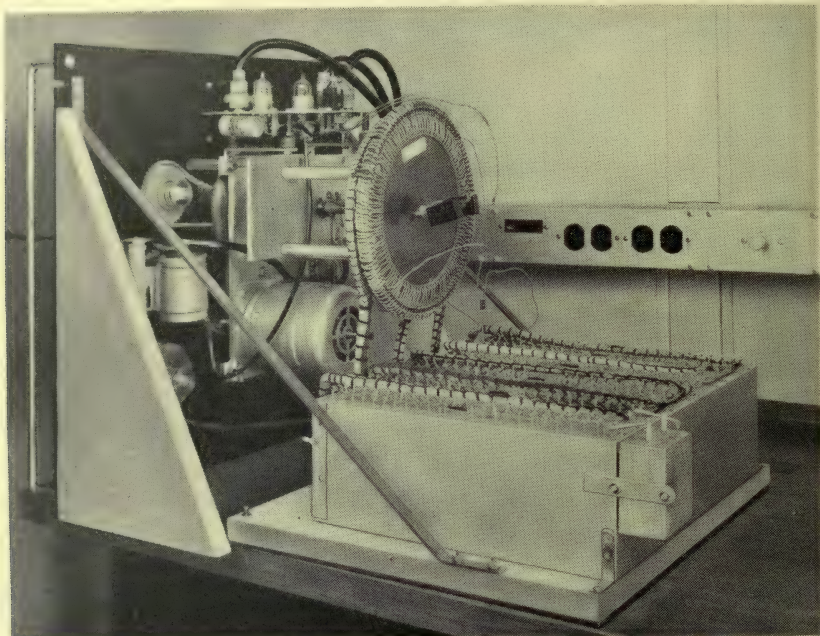


Fig. 6 — Rear view of correlatograph showing rotating switch and delay line.

PRELIMINARY RESULTS

Tests on the complete correlatograph have only been carried far enough to date to verify that the operation is as planned. Fig. 7 shows correlatograms obtained with purely sinusoidal inputs of frequencies 200, 400, 600, 800, 2,000, 3,000, and 4,000 cps recorded on the tape at 15 inches per second. In terms of the original signal the values of τ extend from zero to 12 ms. The profiles represented by the light and dark bands should be rectified cosine waves starting with a peak at $\tau = 0$ and repeating at intervals of one half the period of the wave. In the total range of $\tau = 0.012$ sec, a frequency of 200 cps should show 4.8 periods, 4,000 cps would show 96 periods, and in general f cps would show $0.024 f$ periods. The sudden changes in density parallel to the stripes were caused by manual adjustments of the marking amplifier gain.

Fig. 8 shows correlatograms of a 1,000-cps sine wave embedded in various amounts of flat thermal noise extending throughout the entire input band. The sequence from bottom to top is noise alone, signal and noise power equal, signal power down on noise power by 5 db, 10 db, 15 db, and 20 db. A long time correlation analysis would show zero correlations for the noise alone except for small values of τ . Short

term correlation gives a mottled background level. The signal pattern shows through this background quite plainly when the two powers are equal. As the signal is reduced relative to the noise the signal pattern is obscured and seems entirely missing at 20 db below the noise level. This is a limitation based mainly on the integration time of this particular apparatus. It is possible to improve the resolution by using longer integration periods with corresponding sacrifice of ability to detect fast changes in the applied signal. As pointed out by C. B. Feldman, the autocorrelation type of analysis suffers the same sort of limitations in the low signal-to-noise ratio case as filtering after detection imposes in spectral analysis. That is, finite integration time in the autocorrelation case and finite filter band width in the postdetection filter both allow errors from interaction of noise with noise to swamp the relatively small desired effect of signal.* Cross-correlation on the other hand is like filtering ahead of the detector in that the interaction of noise with noise is suppressed leaving as the dominant error the relatively small interaction of noise with signal. We cannot obtain this advantage of cross-correlation if we do not have available a noise-free signal to cross-correlate with, and analogously an accurate knowledge of the frequency to be selected is helpful in securing the best possible results from predetection filtering.

It will be noted that alternate stripes of the signal pattern fade at first as the signal is decreased relative to the noise. This effect appears to be associated with incomplete suppression by the multiplier of the components proportional to the squares of the individual noise inputs. If the noise inputs were steady, the squares would produce mainly direct current which is not transmitted by the product amplifier. Variation in

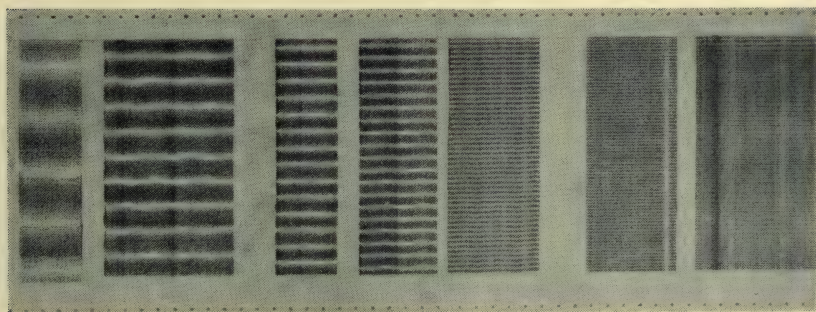


Fig. 7 — Correlatogram of single-frequency wave.

* The autocorrelation method however has an advantage if the signal can be recognized at all in that it is capable of measuring the frequency of the original components, which the postdetection filter cannot do at best.

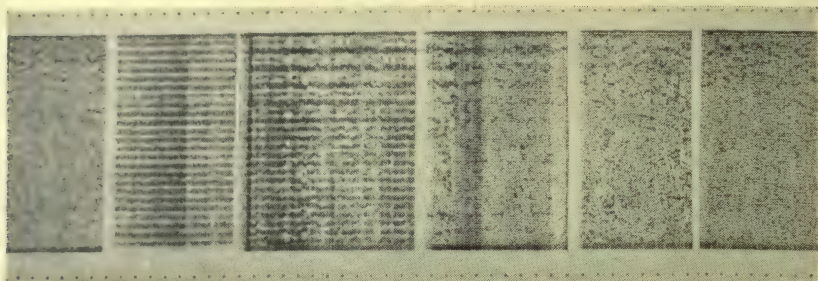


Fig. 8 — Correlatogram of single-frequency wave and random noise.

one noise input because of the idle interval during which no signal is supplied by the delay line is a source of change in the direct current which is partially transmitted through the product amplifier to give a biased integrated value and hence an unequal treatment of positive and negative correlation. Effects of this sort would be particularly noticeable when the noise is large relative to the sinusoidal component.

Fig. 9 shows a sample correlatogram of the sentence "He beats his head against the posts." spoken by G. E. Peterson. The locations of the sounds were marked on the tape by observation during an audio playback and from these marks the corresponding positions on the correlatogram were found. The lower legend gives the ordinary English letters and the upper the symbols of the international phonetic alphabet. The characteristic frequency indicated by the vowel sounds is of the order of 600 cps, which coincides very well with the first formant frequency of the speaker's voice. It is a characteristic of this method that the pattern is dominated by the largest component present. To show the higher formants, which have weaker amplitudes, it would be necessary

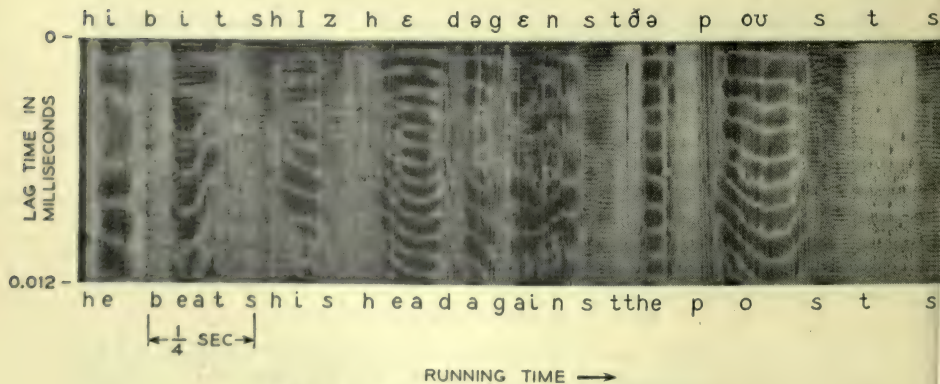


Fig. 9 — Correlatogram of speech sentence.

to filter out the strong low frequency components from the input. The "s" sounds show closely spaced stripes indicating a concentration of energy at the top of the band. Some of the consonants do not show much under the conditions of this test.

It is of interest to compare correlatograms with spectrograms of the same signal. For the single frequency input, the stripes on the correlatogram would be replaced by a single line on the spectrogram. It would be possible to construct signals such that these patterns are interchanged. For example, a rounded band of noise transmitted over two paths having different times of transmission would have a correlatogram with two stripes corresponding to lag times of zero and the delay difference while the spectrogram would show a periodic array of stripes corresponding to the interference pattern of the two paths. It appears that there may be complementary fields of usefulness for the two kinds of analysis and further study is planned.

Besides the many individuals previously named as contributing the various phases of the project, I would like to acknowledge the inspiration and guidance of R. K. Potter in initiating and carrying through the program. L. C. Peterson shared equal responsibility with the author in the project and but for his untimely death would have been a co-author of the present paper.

November 6, 1952

REFERENCES

1. W. Koenig, H. K. Dunn and L. Y. Lacy, The Sound Spectrograph, *J. Acoust. Soc. Am.*, **17**, pp. 19-29, 1946.
2. Y. W. Lee, T. P. Cheatham, Jr. and J. B. Wiesner, Applications of Correlation Analysis to the Detection of Periodic Signals in Noise, *Proc. I.R.E.*, **38**, pp. 1165-1171, 1950.
3. K. N. Stevens, Autocorrelation of Speech Sounds, *J. Acoust. Soc. Am.*, **22**, pp. 769-771, 1950.

A Statistical Study of Selective Fading of Super-High Frequency Radio Signals

By R. L. KAYLOR

(Manuscript received May 7, 1953)

The results of two months of comprehensive frequency-sweep measurements of selective fading in the band between 3750 and 4190 mc over a radio relay path in Iowa are reported. An abridgement of the data, general conclusions derived from the data and an example of the use of the data in connection with frequency diversity measures for radio relay systems are given.

INTRODUCTION

It is well known that, in the high-frequency range during fading conditions, radio signals on different frequencies may exhibit at any instant radically different behavior. This may be true even though they are in the same frequency band and exhibit the same statistical behavior, when observed over a longer period of time. It is also known that h-f fading may be frequency selective enough within the narrow limits of a single radio channel to cause severe distortion of modulated signals. It has been established that the cause of these phenomena is multipath transmission. This knowledge, which is of long standing in the high-frequency range, raised questions concerning the prevalence of similar phenomena in the super-high-frequency range about which relatively little has been known until recently.

During recent years studies of super-high-frequency propagation and fading have been made which have been previously reported.¹ In these tests a frequency-sweep method was used to determine how the loss of a particular radio path varied with frequency at a given instant; and short-pulse methods were used to determine the path length differences which were involved when multipath transmission occurred. These are

¹ A. B. Crawford and W. C. Jakes, Jr., Selective Fading of Microwaves, and O. E. DeLange, Propagation Studies at Microwave Frequencies by Means of Very Short Pulses, Bell System Tech J., **31**, Jan., 1952.

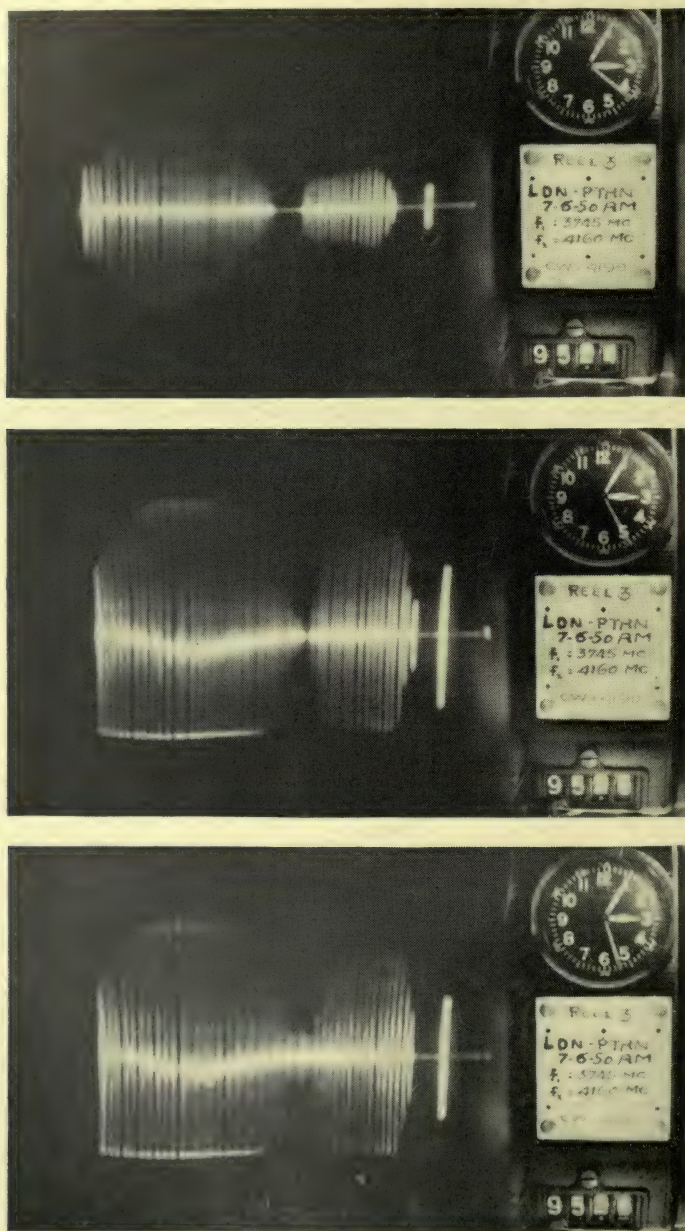


Fig. 1 — Typical records obtained during the Lowden-Princeton, Iowa, frequency-sweep transmission measurements. The horizontal scale is proportional to frequencies between 3,750 and 4,150 mc (the single spike on the right is at 4,190 mc); and the vertical scale is proportional to the amplitude of the received signal.

standard methods for studying multipath transmission (or selective fading) and the results obtained were in good agreement.

Some frequency-sweep transmission measurements were made during the summer of 1950 over a typical radio relay path in the mid-continent region of the United States for the purpose of obtaining additional data of a statistical nature which could be used in system engineering.

The objective of the tests was to determine what per cent of the fading was frequency selective and the degree of its frequency selectivity. Such data were needed to evaluate the advantages of using frequency diversity as a means of minimizing the effects of fading, and to provide quantitative data for designing frequency-diversity measures. Recordings were made throughout the months of July and August which serve as the basis for a statistical picture of the fading occurring during those months in the frequency band between 3,750 and 4,190 mc.

Path Over Which Measurements Were Made

The measurements were made over a 30.8-mile path between the Lowden and Princeton (Iowa) towers of the Chicago-Omaha Radio Relay System. This path was chosen as a typical radio relay path as to length, clearance above terrain, and climatic conditions. Height-loss runs made by the American Telephone and Telegraph Company (when originally selecting the path as part of their radio relay route) indicated that ground reflections on this path were unimportant. The reflection coefficient was less than 0.1

Method of Tests

The frequency-sweep equipment used in these tests was similar to that used in the previous studies.¹ With this equipment about 50,000 record photographs were obtained of a cathode-ray tube presentation of the path-loss vs frequency characteristic of the path between 3,750 and 4,150 mc. Fig. 1 shows three illustrations of these record photographs, which are more fully discussed below. The taking of these records was distributed throughout the months of July and August in such a manner as to give complete coverage of all the fading during that period. In addition the single-frequency path loss at 4,190 mc was continuously recorded throughout the two months.

The data from these records were analyzed on a statistical basis; and families of curves were obtained depicting several aspects of the nature of the selective fading encountered during the tests. These are described more fully below.

GENERAL DISCUSSION OF FADING PHENOMENA

The variations in the strength of a received radio signal known as "fading" are caused by variable or temporary conditions in the transmission path. These conditions fall into two broad classes: those causing partial obstruction of the path, and those causing multipath transmission. The latter class is believed to be the principal source of frequency-selective fading.

Multipath transmission involves the reception of more than one signal ray, each of which travels over a different path between the transmitter and receiver. Generally each such path has a different length between the transmitter and receiver. Multipath transmission involves either or both reflection or refraction of at least one (and in some cases all) of the received rays. Since the conditions of the atmosphere are continuously varying, the paths of the received rays are variable with time. The received signal is the resultant of all the rays accepted by the receiving antenna. The relative phase of the different received rays depends on (1) the differences in the lengths of the paths over which they have travelled and (2) the signal frequency. If the rays arrive at the receiving antenna in nearly the same phase, they add and enhance the received signal; if they arrive in phase opposition they partially cancel each other and fading results. This fading is not only variable with time but also with signal frequency, and is called "(frequency) selective fading". Since the conditions which cause this kind of fading are substantially random, the variation of fading with time on a statistical basis might be expected to approach the Rayleigh distribution. There has been experimental confirmation of this.

A number of other factors are involved which will not be treated here, since the mechanism and effects of multipath transmission have been discussed quite thoroughly in the previously mentioned reports by Crawford and Jakes, DeLange and in an earlier paper.²

Results of Tests

Fig. 1 shows three illustrations of the type of records obtained. On each record the horizontal deflection of the cathode-ray-tube trace is proportional to the radio signal frequency. The vertical deflection is linearly proportional to the amplitude of the received signal, which because of constant transmitter power is inversely proportional to the

² H. T. Friis, Microwave Repeater Research, Bell System Tech. J., 27, Apr., 1948.

path loss.³ In each record one second was required for the trace to travel through the entire frequency range covered.

The three records shown are consecutive, being taken at 3:05:21 AM, 3:05:25 AM and 3:05:27 AM. The first record (frame number 9524) was taken with 10 db more attenuation in the input to the measuring equipment than when the later two records were taken (frame numbers 9525 and 9526). There is obvious overloading in the left hand portions of the records on frames 9525 and 9526; but the accuracy of the right hand portions of these records (showing the deeper part of the fade) is unimpaired. The noticeable difference between the shapes of the curves near the deeper parts of the fade on frames 9525 and 9526 is typical. These changes occurred within two seconds. Generally, the deeper parts of the fades show more rapid changes than the less deep parts.

From the 50,000 record photographs all those pertaining to fading of 30 db or more were segregated and analyzed. The remaining records were analyzed on a sampling basis, except that every record showing unique effects was analyzed. About 1,800 path-loss versus frequency curves such as those illustrated in Figs. 2 and 3 were obtained. These curves were separately studied, and were also treated statistically.

Because the levels during the deeper part of the fade are too low to show on frame 9524, and because portions of frame 9525 showing the less severe part of the fade are affected by overloading, it was necessary to combine two records (shown on Fig. 1) to obtain the single path-loss versus frequency curve shown as Fig. 2(a).

Theoretically, it should be possible to synthesize by means of an addition-of-vectors method each of the path-loss versus frequency curves obtained from these tests. Each vector term of the equation would correspond to a component of the received signal and would be of the form $R \cos \omega T$, where: R is the magnitude of a particular component (normalized to the magnitude of the direct signal component), T is the delay (in seconds) between the time of arrival of the direct signal component and the particular interfering component, and ω is 2π times the frequency.

In practice, however, it has been found in the case of deep fades that components of relatively small magnitude and relatively long delay are of importance in determining the shapes of the curves near maxima of path loss. Also it has been found that in most such cases quite a few components are involved. These factors make accurate analysis of many

³ A logarithmic amplitude characteristic (db scale) would have been preferable; but time did not permit modifications of the test equipment before the start of the 1950 fading season.

curves impractical without a computer capable of handling a relatively large number of components several of which may be relatively small.

However, experimentation with a trial and error method of solution on a few selected curves has shed considerable light on the nature of received signals which could produce the types of curves included in the data under consideration.

Curve (a) on Fig. 4 is one of those selected for analysis and was synthesized by using a direct signal component and six interfering components as shown in Table I. If components 5 and 6 had been omitted from the synthesis, curve (b) would have resulted; and, if components, 3, 4, 5 and 6 had been omitted curve (c) would have been obtained.

The difference between curves (a) and (b) illustrate the radical effect which can be produced on the shape of a curve near maxima of path loss

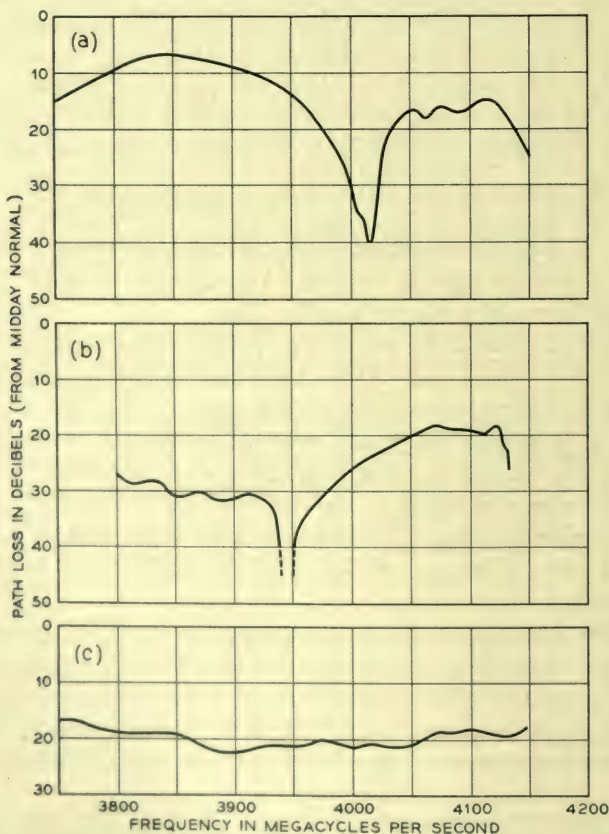


Fig. 2 — Typical path-loss versus frequency curves observed on the Lowden-Princeton, Iowa, path during July and August, 1950.

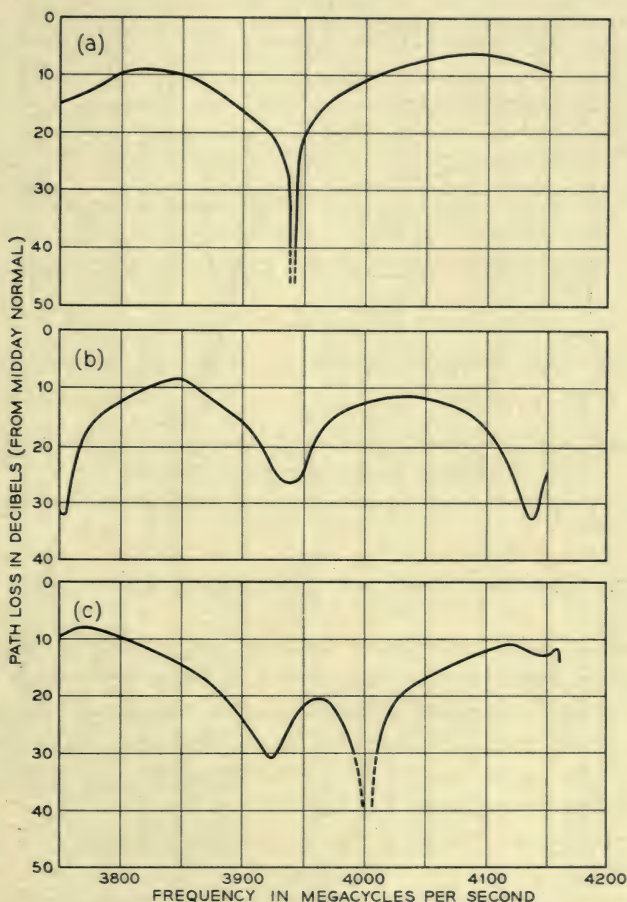


Fig. 3 — Typical path-loss versus frequency curves observed on the Princeton-Lowden, Iowa, path during July and August, 1950.

by relatively small components such as components 5 and 6. In the case of deeper fades much smaller components may be of importance in determining the shapes of the curves near maxima of path loss; hence the curves for deep fades are quite difficult to synthesize. The shape near the maximum of path-loss on the curve in Fig. 2(a) is quite typical of the shapes to be found in the data under discussion.

Based on the experience gained in attempting to synthesize some of the curves, it is possible to recognize the significance of certain features of other curves, and to generalize the nature of all the curves. The principal conclusions are:

1. No deep fades were found which did not show definite frequency

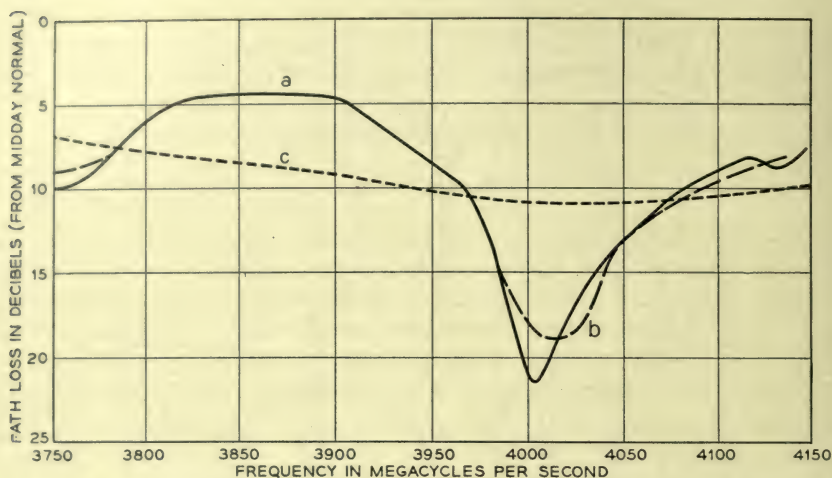


Fig. 4. — Path-loss versus frequency curves synthesized from components shown in Table I.

selectivity. (Statistical data are given later which bear out this conclusion).

2. No deep fades were found that did not include an appreciable component of loss with little appearance of frequency selectivity; that is, no deep fading occurred unless the signal was already depressed 10 or so db across the entire observed band. This could be caused either by (a) the presence of appreciable interfering signals of slight delay from the direct signal component, see curve (c) on Fig. 4, or (b) by non-frequency selective attenuation of the direct signal component. The former seems more likely; but the evidence is not conclusive in this regard.

3. No curves for deep fades were found which appeared as if they could be synthesized satisfactorily with fewer than four to six components.

TABLE I

Component	R (Normalized to magnitude of direct received signal)	T Milli-Microseconds	Phase Shift From Direct Signal Component	
			Half Wave Lengths	At Frequency
0	1.0	0	0	Any Freq.
1	0.45	0.122	1	4090 mc
2	0.26	0.370	3	4050 mc
3	0.10	2.86	23	4033 mc
4	0.115	3.14	25	3981 mc
5	0.025	3.9	31	3972 mc
6	0.02	12.1	97	3993 mc

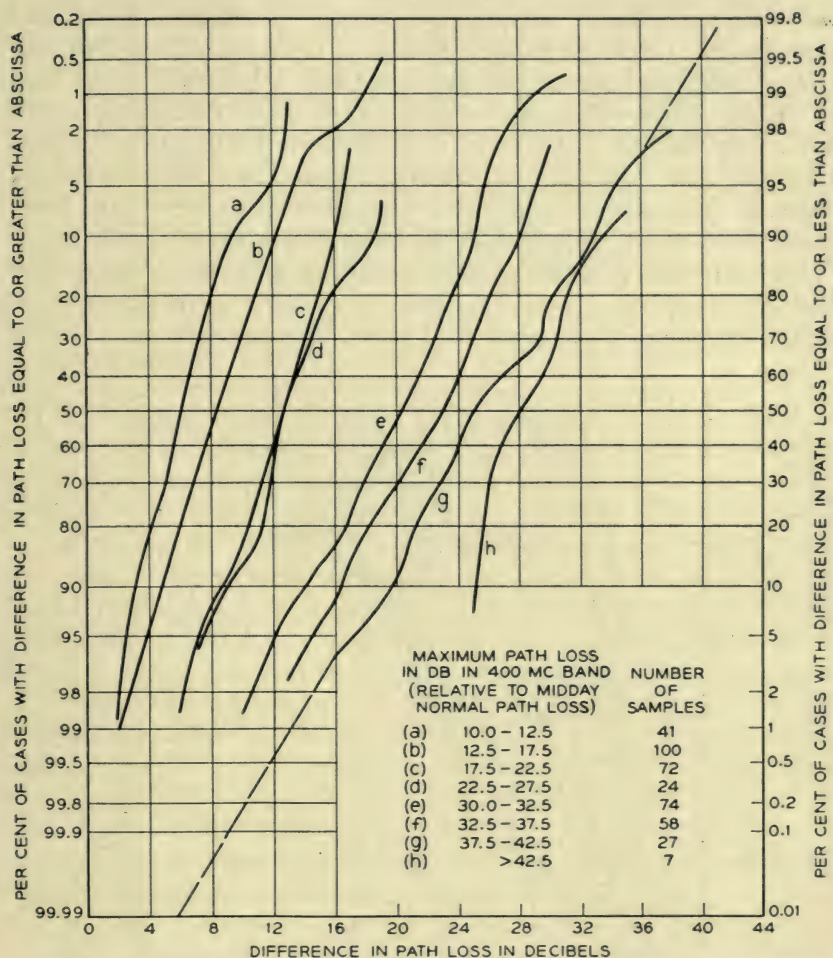


Fig. 5 — Statistical distribution of differences between maximum and minimum path loss in bands 400 mc wide. (Center frequency of 400-mc band random with respect to frequencies corresponding to maxima and minima of path loss).

4. It seemed to be the rule that components with relatively long delays were appreciably smaller in magnitude than those with shorter delays.

5. At the bottom of the deep fades some of the small signal components with relatively long delays were an important factor in determining the shape of the path-loss versus frequency curves.

Fig. 5 illustrates one type of statistical information which was derived from the 1,800 individual path-loss versus frequency curves. The family of curves on Fig. 5 shows the per cent of fades (of a given depth) which

were frequency selective to any given degree. For example, 95 per cent of fades which were 40 db deep showed a variation in path loss (frequency selectivity) across the observed band 400 mc wide of at least 17 db. By extrapolating the curve for 40 db fades (which appears to follow a normal distribution law) it can be estimated that 99 per cent of the 40 db fades had at least 13 db of frequency selectivity within the observed band 400 mc wide. Probably, observation of a wider band would have shown even greater evidence of frequency selectivity.

During periods when the received signal was materially stronger than mid-day normal there was practically no evidence of frequency selectivity. Several periods were observed when the path-loss across the entire

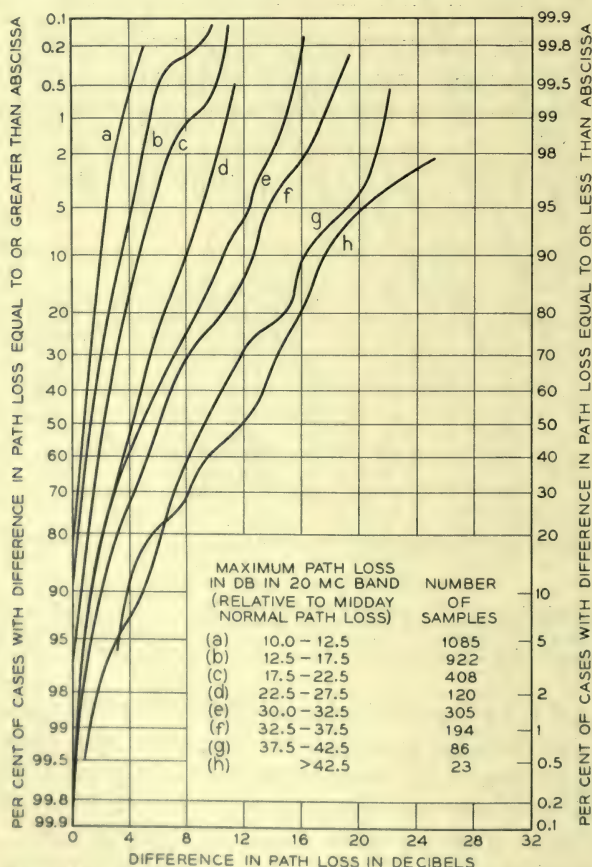


Fig. 6 — Statistical distribution of differences between maximum and minimum path loss in bands 20 mc wide. (Center frequency of 20-mc band random with respect to frequencies corresponding to maxima and minima of path loss).

400-mc observed band was 8 db less than mid-day normal; and in one case it was more than 10 db less than mid-day normal.

Another family of curves is shown by Fig. 6. These curves were obtained by dividing the 400-mc observed band into a number of 20-mc bands, chosen at random with regard to the shapes of the path-loss versus frequency curves. These data show the difference between the maximum and minimum losses within a single 20-mc broad-band channel which might accompany a fade of a given depth. Such data are of use in estimating possible distortions of a modulated signal occupying a band width of 20 mc.

Frequency Diversity

The fact that the instantaneous fading may be different on different frequencies within the same frequency range offers a means for mitigating transmission impairments caused by fading. During periods when there is fading in excess of a specified value on the regular carrier frequency, the carrier can be shifted to an alternate frequency in the hope that the fading on the alternate frequency may be less severe. The merit of using this type of frequency diversity can be gauged from the statistical distribution of fading on the alternate frequency during periods when there is fading in excess of the specified value on the regular frequency.

The data obtained in these tests indicates that there was no correlation between the fading on frequencies separated by: (1) 40 mc or more during periods when there was fading of 10 db on one of the frequencies and (2) 160 mc or more during periods when there was fading of at least 20 db on one of the frequencies. However, during periods of severe fading, the data indicate considerable correlation between the fading on regular and alternate frequencies separated by 80 mc or less.

Fig. 7 shows the distributions of fading on alternate frequencies at specified frequency separations from the regular frequency, when there is fading of a specified depth on the regular frequency. Table II indicates which of the curves on Fig. 7 applies to a particular set of conditions.

Curves *G* and *H* on Fig. 8 show the distributions of depths of fade at 4190 mc during the entire months of July and August, respectively. Comparison of the 4,190-mc data with data from tests made over other paths indicates that the fading occurring during the Iowa tests was nearly as severe as during the "worst month" observed on any path to date.

Curve *B* on Fig. 7 shows the statistical distribution of fading on any frequency (within the range under consideration) during periods when important fading is prevalent. The fading shown by this curve is con-

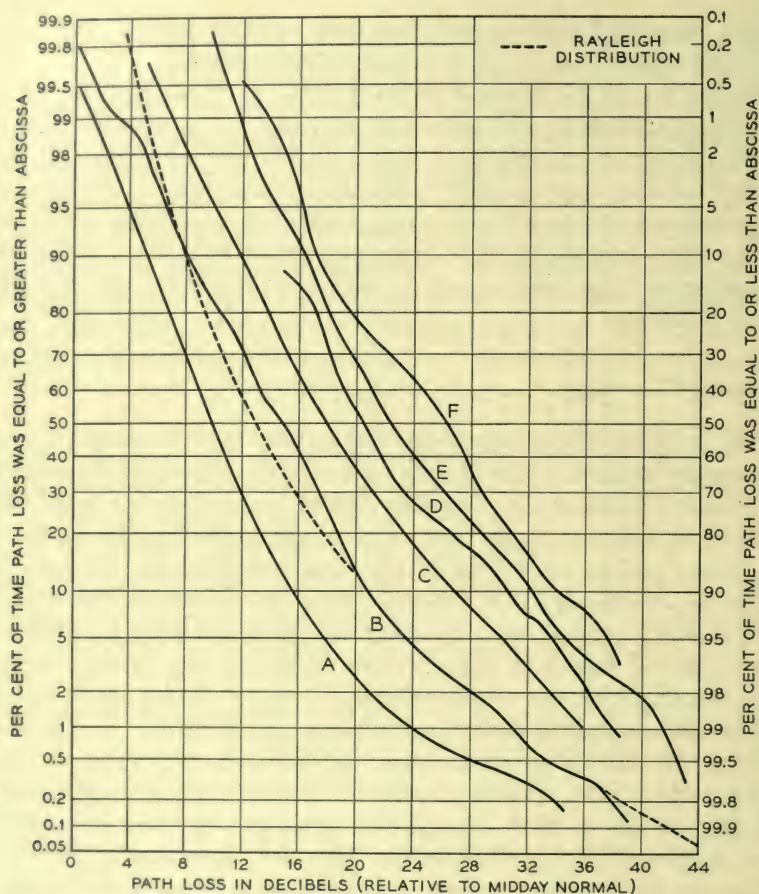


Fig. 7 — Statistical distributions of fading measured on Lowden-Princeton, Iowa, path during July and August, 1950. These curves are useful in predicting the efficacy of specific frequency-diversity systems as described in the text.

siderably more severe than is shown by the curves on Fig. 8. This is because the latter curves include all the time within the months under consideration. Therefore, they include much time when the fading mechanism was not present in the path, and the path loss remained steady near mid-day normal.

There is evident similarity between the shapes of Curves *A* through *F* (on Fig. 7) with the shape of the curve which is based upon the Rayleigh distribution. This is consistent with the theory that deep fading associated with multipath transmission is caused by random phasing of a large number of vectorial components.

These data have been applied practically to the design of frequency-diversity measures for minimizing circuit outages caused by fading in TD-2 radio relay systems.

Fig. 9 shows an illustration of the practical use of the data. Curve *H* (which is the same as Curve *H* on Fig. 8) shows the distribution with time of the fading on a typical radio relay path without frequency diversity measures. Curve *J* shows the distributions of fading if certain specific frequency diversity measures are used (based solely on fading considerations). The difference between Curves *H* and *J* shows the improvement gained by using that specific kind of frequency diversity. For example, Curve *H* shows that without frequency diversity fading in excess of 30 db will occur 0.075 per cent of the time and fading in excess of 40 db will occur 0.02 per cent of the time. But, if the kind of fre-

TABLE II

Depth of Fade on Regular Frequency	Separation Between Regular and Alternate Frequencies	Curve
10 db	40 mc or more	A
20	40 mc	C
20	80 mc or more	B
30	40 mc	E
30	80 mc	C
30	160 mc or more	B
40	40 mc	F
40	80	D
40	160 mc or more	B

quency diversity to which Curve *J* corresponds is used, fades deeper than 30 db will occur only 0.0012 per cent of the time, and fades deeper than 40 db will occur only 0.00008 per cent of the time. Thus the improvement resulting from this type of frequency diversity is a reduction of fades deeper than 30 db from 0.075 to 0.0012 per cent of the time, and fades deeper than 40 db from 0.02 to 0.00008 per cent of the time.

To explain how Curve *J* was derived, let us assume that during any time when there is a fade of 30 db or more on the operating frequency of a given channel the signal will be switched to another channel frequency. Let us further assume that the alternate frequency will be separated by 160, 240, 320, or 400 mc from the assumed operating frequency and that the choice of alternate frequency is random. If we also assume that the conditions of August, 1950, on the Iowa path prevail, we will find from Curve *H* on Fig. 8 that the assumed operating frequency will have a fade of 30 db or more 0.075 per cent of the time. Then a new distribution curve can be prepared, based on (1) Curve *B* (on Fig. 7) for

0.075 per cent of the time, and (2) that portion of Curve *H* (on Fig. 8) which corresponds to fades of less than 30 db for the remaining 99.925 per cent of the time. Curve *J* on Fig. 9 is such a curve.

If frequencies separated from the assumed operating frequency by 80 mc were used instead of those assumed above, Curve *C* instead of Curve *B* on Fig. 7 would have been used. This would have shown a somewhat smaller improvement because of some correlation between the fading on the assumed operating frequency and an alternate frequency separated from it by 80 mc. However, the improvement gained from the use of this system would be ample to prove-in its use.

The question is sometimes raised concerning the reason that the curves on Fig. 9 show an apparent improvement for fading of less than 30 db, since the circuit is switched only when fading deeper than 30 db occurs. This is because the curves are cumulative distribution curves;

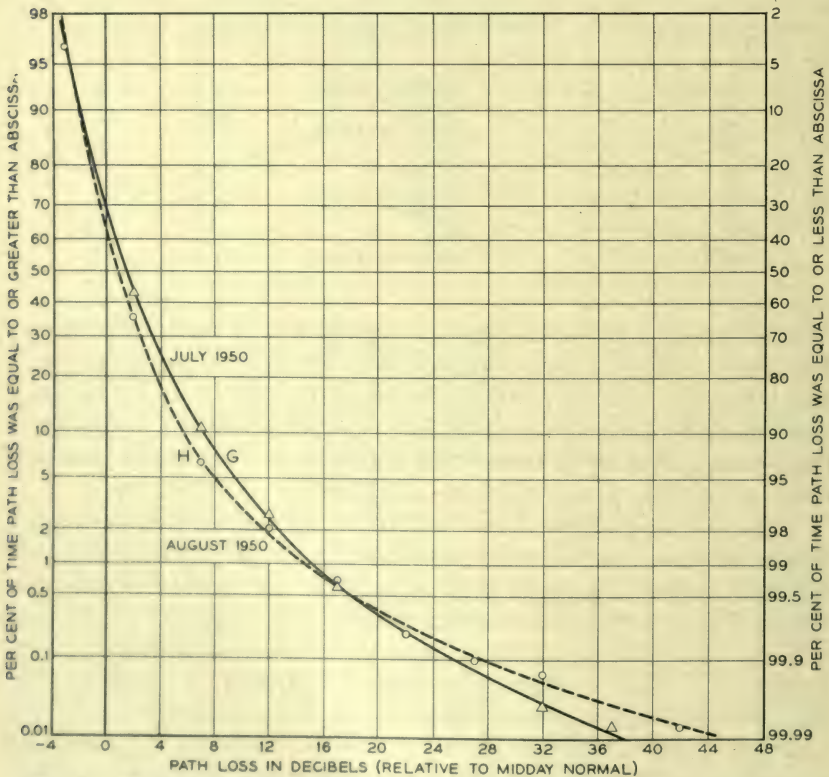


Fig. 8 — Statistical distribution of fading loss Princeton-Lowden, Iowa, path July and August, 1950.

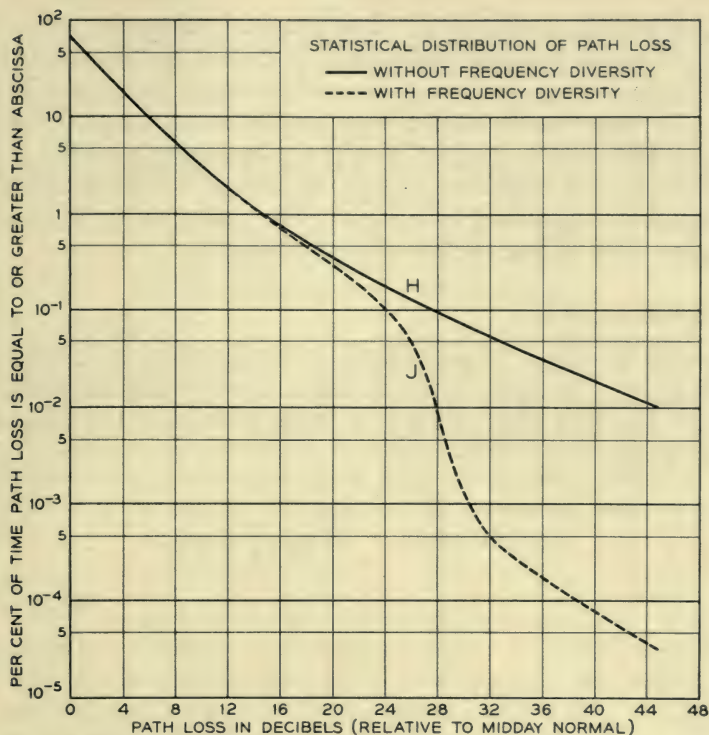


Fig. 9—Effect on transmission of use of frequency diversity in 3,750–4,150 mc range.

and reduction of the per cent of time when there is fading deeper than 30 db also reduces the per cent of time when there is fading deeper than 25 db, etc.

CONCLUSIONS

Quantitative information on fading phenomena is essential in the engineering of super-high-frequency systems and in evaluation of frequency diversity arrangements. The data obtained from the field tests and the statistical analysis reported herein, while limited to a single path and particular season, fill a gap in previous knowledge. They have found practical application in the design of diversity systems for improving the reliability of radio relay systems.

The principal conclusions that can be drawn from the available data are: (1) all of the deep fading is definitely frequency selective, and is caused by a complex multi-path transmission; (2) deep selective fading

is ordinarily accompanied by a 6 to 10 db signal depression over a band of at least several hundred megacycles; (3) fading on frequencies separated by 160 or more megacycles shows little correlation, but fading on frequencies more closely spaced shows an increasing correlation as the frequency difference is diminished; and (4) frequency diversity offers a practical means of mitigating circuit impairment due to fading, if the transmission can be shifted to a frequency far enough away to minimize correlation with fading on the original frequency.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the important contribution of J. Mallett, who conducted the field tests during which the data were obtained. The field tests and data analysis were under the supervision of R. P. Booth.

Acceleration Effects on Electron Tubes

By F. W. STUBNER

(Manuscript received June 19, 1953)

This paper discusses methods of measuring shock and vibratory accelerations to which electron tubes may be subjected in various equipments, and the influence of these disturbances on the performance of the tubes. An outline is given of the design problems connected with the elimination of tube damage or faulty operation of tubes under adverse shock or vibration conditions, and of the methods used for simulating these conditions by means of production test machines and test methods.

INTRODUCTION

The rapid expansion of the use of electronic equipment by industry and the armed services has created increasingly new demands on both the electrical and mechanical characteristics of electron tubes. Since these tubes are electronic devices, it is only natural that their structural designs are dictated to a large extent by electrical requirements. However, the experience gained with conventional tubes in some of the new equipment applications has revealed certain mechanical shortcomings which reflect on the proper functioning or life of the tubes. The realization of the increasing importance of mechanical design has resulted in an increased effort for structural improvements to assure more reliable tube performance. The term "reliable", as used here, denotes that a tube has a high degree of dependability when subjected to specific conditions, either electrical or mechanical. Thus, the requisites for reliability may differ for various applications. Although, in designing a tube, many requirements must be taken into consideration, only some of the problems connected with dependable tube performance under mechanical disturbances, i.e., shocks and vibrations, will be discussed here.

Since the performance of a tube depends on the geometry of its component parts, minute changes in element spacing may produce variations in its characteristics. Because of the necessarily delicate structure of some tube elements, permanent or transient dimensional changes may be produced by mechanical forces acting on the tube unless the tube is

specifically designed to minimize the effects of such disturbances. For a rational design, it is, therefore, necessary to have some knowledge of the nature of the disturbances likely to be encountered by tubes during their service life. If one considers the numerous conditions under which electronic equipment is required to function, it becomes clear that the mechanical requirements are manifold indeed. Equipment applications can be divided into three general groups, each one imposing in many cases special requirements on tubes. These groups are: (a) stationary equipment, such as central office telephone installations, home radio and television receivers, etc., (b) mobile equipment used in land vehicles, ships or airplanes, and (c) portable equipment. In many instances, military equipments straddle above subdivisions and superimpose additional requirements. It must be stressed, too, that pre-service conditions encountered through handling and shipping must be taken into consideration, since a tube is of no use to the customer until it is installed and operating.

A knowledge of expected service conditions is not only useful in the initial design stage but also aids the manufacturer in devising suitable test gear to check the quality of the product at the factory. Although a wealth of data has been collected on shocks and vibrations to which electronic equipments are subjected under actual service conditions, little is known how these disturbances are altered by the mechanical structures of the equipments before they reach the tubes. In general, the nature and magnitude of mechanical disturbances can be expressed in terms of acceleration, velocity, or displacement. Since electron tubes may respond to a wide frequency range of vibrations, the most sensitive measure is acceleration, which varies as the square of the frequency of the element displacement. Velocity or displacement instruments are usually not sufficiently sensitive to give a true record of disturbances in the higher modes of vibration due to the small velocity and displacement values involved.

The investigation of the nature of accelerations at tube sockets offers special problems. The accelerometers must approximate the weights of the tubes used in the respective sockets so that the disturbances are not modified by the substitution of acceleration pick-ups for the tubes. For the same reason, the method of fastening the accelerometers in the sockets must duplicate that of the tubes, and since the accelerating forces may act in several directions, the accelerometers must be capable of exploring these various directions. Lightweight accelerometers meeting the above requirements have been developed recently. These instruments are generally built to approximate the weight, weight distribution, and shape of vacuum tubes. A more detailed description of these accelerometers and associated recording circuits is given in the following chapter.

ACCELEROMETERS AND ASSOCIATED INSTRUMENTATION

Instrumentation employed for measuring accelerations depends largely on the frequency range of the disturbances to be recorded, which, for vacuum tube applications, covers approximately the audio spectrum.

In practically all cases, electronic recording equipment is used because of its high degree of flexibility. Its basic components consist of (a) a mechanical-electrical transducer which translates mechanical disturbances into proportional electrical potentials, (b) an electronic amplifier to step up the voltage output of the transducer to desired levels, and (c) an indicating instrument, usually a cathode ray oscilloscope, for observations of the disturbances.

Accelerometers

Although several types of mechanical-electrical transducers or pickups are in existence,¹ it has been found that the self generating types employing materials such as quartz or ferro electric crystals, are most useful since they can be constructed so that their output is directly proportional to acceleration over a large frequency range. Velocity and displacement indicating instruments are sometimes used for low frequency work; by suitable differentiation their output signal will be proportional to acceleration.

An accelerometer must have the following fundamental properties:

- (a) the signal produced must be proportional to the accelerations to be measured.
- (b) its calibration must be stable and unaffected by humidity and temperature changes encountered.
- (c) its mechanical strength has to be adequate to withstand the acceleration stresses to which it is subjected.
- (d) its weight must be sufficiently low so that the disturbance patterns are not modified by loading.
- (e) its sensitivity (voltage output/g) and useful frequency range must be sufficiently high to cover acceleration magnitudes and frequency components to be recorded.

A number of piezoelectric materials are available and have been employed in various electro-mechanical transducers. The development of practical light-weight transducers has been made possible by the use of some of the relatively new ferro-electric ceramics. By proper compounding, firing and poling, these materials can be made to have very high sensitivity and good life stability. Other advantages are their high dielectric values, good mechanical strength and relative insensitivity to usually

encountered humidity and temperature conditions. Accelerometers approximating the size and weight of electron tubes are now commercially available and used for measuring accelerations imparted to tubes through their sockets.

By proper construction the voltage output of these pick-ups is proportional to the stresses induced in the active elements within the desired frequency range. Depending on choice of material and construction, their active elements can be used in compression, shear, or torsion. Fig. 1(a) shows the simplified construction of a compression type pick-up. It consists of (1) a base which is rigidly fastened to the point at which accelerations are to be measured, (2) a sensitive element, (3) a weight, and (4) a retaining spring. The whole assembly is shown held together by a screw. The mechanical equivalent of this structure is shown in Fig. 1(b).

When this unit is subjected to sinusoidal motion its displacement is given by

$$X = X_0 \cdot \sin \omega \cdot t$$

Where X = instantaneous displacement of the base from equilibrium at time t

X_0 = maximum displacement from equilibrium

ω = circular frequency of the motion

The instantaneous acceleration then becomes

$$a = \frac{d^2x}{dt^2} = -X_0\omega^2 \sin \omega t \quad (1)$$

The motion is transmitted to the mass M through the parallel spring

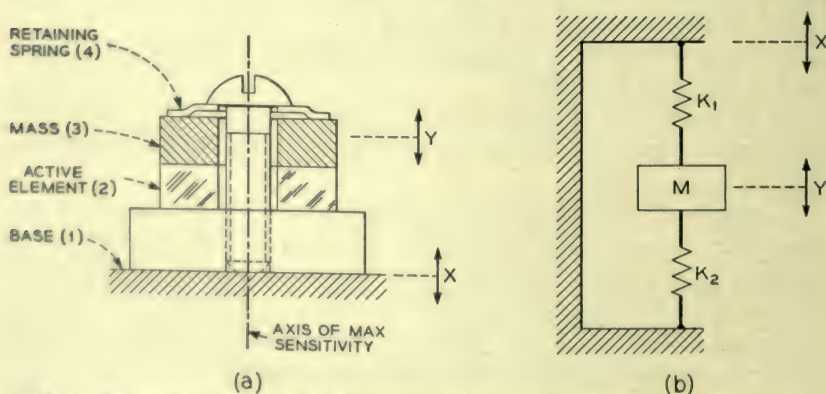


Fig. 1 — Compression type accelerometer (a) and its mechanical equivalent (b).

system formed by the stiff active element K_2 and the retaining spring K_1 . For negligible damping in these springs the motion of the mass relative to the base is:

$$y_1 = \frac{X_0 \left(\frac{\omega}{\omega_n} \right)^2}{1 - \left(\frac{\omega}{\omega_n} \right)^2} \cdot \sin \omega t \quad (2)$$

Here ω_n = the natural circular frequency of the mass on the springs. The force F exerted by the springs on the mass is:

$$F = (K_1 + K_2)y_1 = \frac{(K_1 + K_2)X_0 \left(\frac{\omega}{\omega_n} \right)^2}{1 - \left(\frac{\omega}{\omega_n} \right)^2} \sin \omega t \quad (3)$$

The accelerometer is constructed so that the spring constant K_1 of the retaining spring is considerably smaller than that of the active element K_2 . Therefore, for vibration frequencies relatively low as compared to ω_n , equation (3) becomes:

$$F = \frac{K_2}{\omega_n^2} X_0 \omega^2 \sin \omega t \quad (4)$$

Where

$$\frac{K_2}{\omega_n^2} = M, \quad \text{and} \quad X_0 \omega^2 \sin \omega t = a$$

The stress produced by this force F produces a charge on the active element which is proportional to the instantaneous value of the acceleration to which the unit is subjected. Properly calibrated therefore, the acceleration can be measured in gravitational units. When the disturbing force consists of a number of these components, the total instantaneous output is proportional to the sum of these components, within the frequency limitation of the pick-up.

In deriving the above expressions, a number of assumptions have been made. Equation (4) therefore holds only if:

(a) ω_n is made large compared to the highest shock or vibration frequencies that are to be recorded.

(b) K_1 is small compared to K_2 . Since the function of the retaining spring is merely to hold the assembly together for peak accelerations encountered, a soft spring with a sufficiently large static deflection can be employed or the spring may be replaced by conducting cement to hold the parts together.

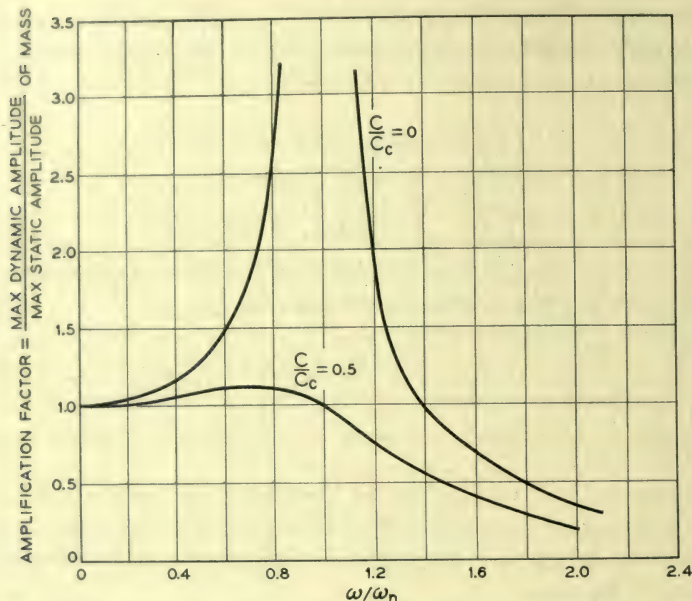


Fig. 2 — Amplification of forced vibration on mass for two degrees of damping.

(c) damping of the system is small. If damping is considerable, equation (2) becomes

$$y_1 = \frac{\left(\frac{\omega}{\omega_n}\right)^2 X_0 \sin \omega t}{\left(\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left[2 \frac{c}{c_c} \frac{\omega}{\omega_n}\right]^2\right)^{1/2}} \quad (5)$$

In this expression c/c_c = fraction of critical damping. However, since by construction, the natural frequency of the system is large compared to the forcing frequencies, and damping of the active elements employed in these accelerometers is small, the term $\left(2 \frac{c}{c_c} \cdot \frac{\omega}{\omega_n}\right)^2$ becomes negligible. Equation (5) therefore simplifies to equation (2).

The useful upper frequency limit is generally fixed at $\frac{1}{4} \omega_n$, because, as shown by Fig. 2* amplification of the impressed signal results as ω/ω_n increases. If the disturbances contain frequencies beyond the useful limit, low pass electrical filters must be employed in the associated amplifier to suppress these frequencies. The fact that this type of pick-up has very low damping offers some disadvantage in recording complex waves be-

* Figs. 2 and 3 were obtained from material given in Reference 2.

cause phase distortion between the impressed and recorded frequency components takes place. Fig. 3* graphically illustrates this relation. This figure shows that $\left(\frac{c}{c_c}\right)$ would have to be made approximately 0.5 in order to have a phase angle proportional to the impressed frequencies, i.e., to obtain a true recorded pattern of the disturbance. Damping is generally not employed in accelerometers built to duplicate the size and weight of electron tubes due to constructional difficulties. The useful frequency limit of these devices is often not governed by the resonant frequencies of the active elements but by the lowest resonant frequency of their housing structure.

As mentioned previously, the active elements of these light weight transducers can be employed in various ways. The first models which were developed a few years ago made use of the elements in compression, following the practice of their predecessors, or: the relatively heavy and large quartz crystal accelerometers. A working model of a compression unit is shown in Fig. 4(a). Although this type of unit can be constructed to have a very wide useful frequency range (high resonant frequencies), it possesses two disadvantages: its internal capacity is rather low, and it has poor directional sensitivity. Its sensitivity decreases to roughly

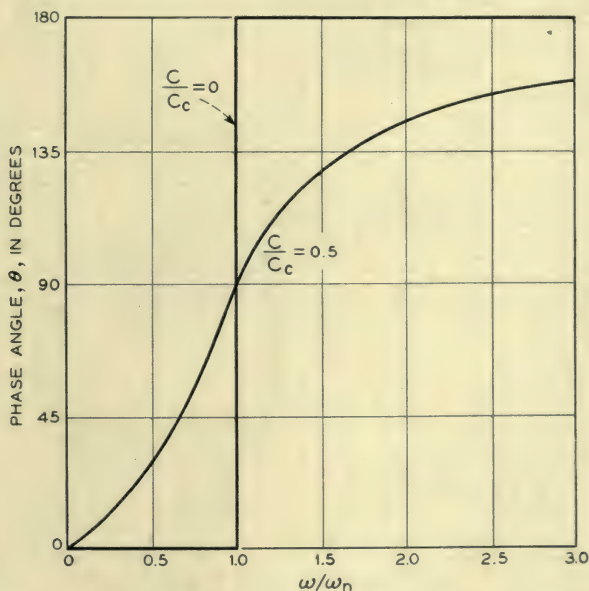


Fig. 3 — Phase angle between force and displacement as a function of the forcing frequency for two values of damping.

10 per cent of maximum in the direction perpendicular to the sensitive axis. Some improvements in directivity were obtained by utilizing the elements in shear. Fig. 4(b) illustrates a tandem shear type accelerometer. The construction of this unit is somewhat more complicated. The elements first have to be poled in the proper direction. The conducting layers used for poling in this direction then have to be removed and new conductive coatings have to be applied on the areas facing the base and mass M . The unit shown is sensitive to accelerations in the radial direction. The improved directivity of the shear elements over compression elements is shown in Fig. 5. Although shear elements are superior for detecting acceleration components along desired directions, it is difficult to make their internal capacity sufficiently high without increasing the over-all weight of the accelerometer for a given sensitivity. The use of cantilevers for these applications also suggested itself, for relatively high tensile and compressive stresses can be produced in such structures. Cantilever type elements have been constructed as shown schematically in Fig. 6(a). Here thin laminations of active material are cemented on opposite sides of a small metal cantilever. Under the action of external forces bending of the inner member subjects the outer laminations to

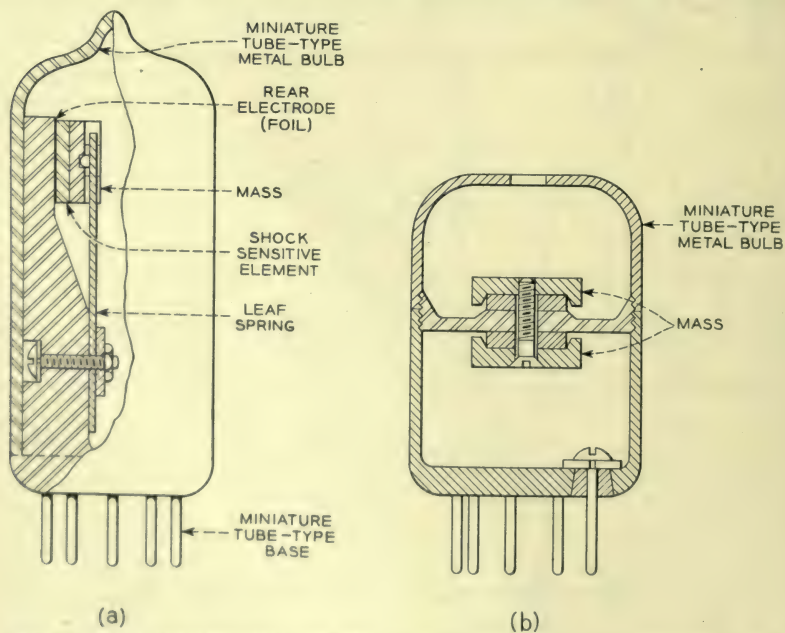


Fig. 4—(a) Radial miniature tube compression type accelerometer and (b) radial miniature tube shear type accelerometer.

tension and compression stresses respectively, thereby producing the desired charges in these elements. The characteristics, i.e., the resonant frequency and sensitivity, of these elements are largely determined by the dimensions of the metal member. Relatively high internal capacities can be obtained by this construction. A recent variation of such a structure is shown in Fig. 6(b). In this design the inner metal component of the cantilever described above has been eliminated. The entire structure is made of the active material, in this case in cylindrical shape for ease of

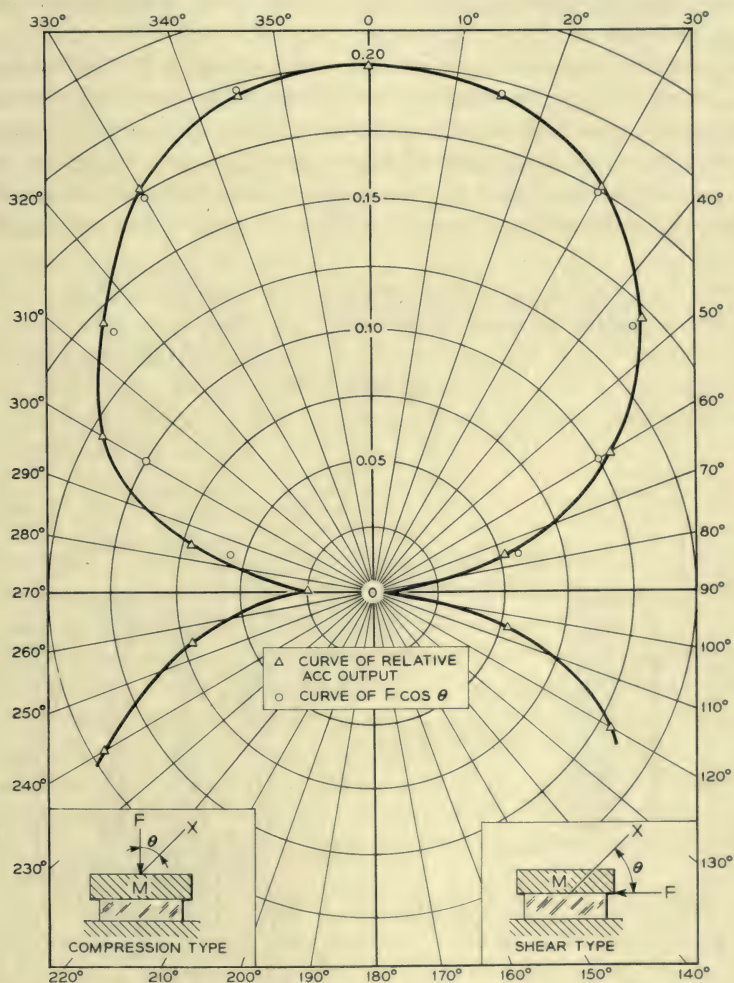


Fig. 5 — Relative angular response of compression and shear type ferro electric crystal accelerometers.

manufacture. Sections of conductive coating are deposited on this cylinder as shown. These are used to initially polarize the material, and, in use, to collect the charges produced by stressing the inner and outer fibers of the cylinder. It is interesting to note that these elements can be used to detect, by proper choice of connections, either radial or axial accelerations. Through suitable external instrumentation both directions can also be recorded simultaneously if desired.

The above illustrations show but a few applications of this material for detecting accelerating forces. Additional forms will suggest themselves, each of particular advantage for specific application.

The sensitivity of an accelerometer is generally given in coulombs/g or open circuit voltage. Its effective voltage output over a given frequency range is a function of the characteristics of the associated equipment including the connecting cable. Calibration in the lower frequency range up to a few hundred cycles may be performed by vibrating the accelerometer on variable frequency vibration machines or resonating spring

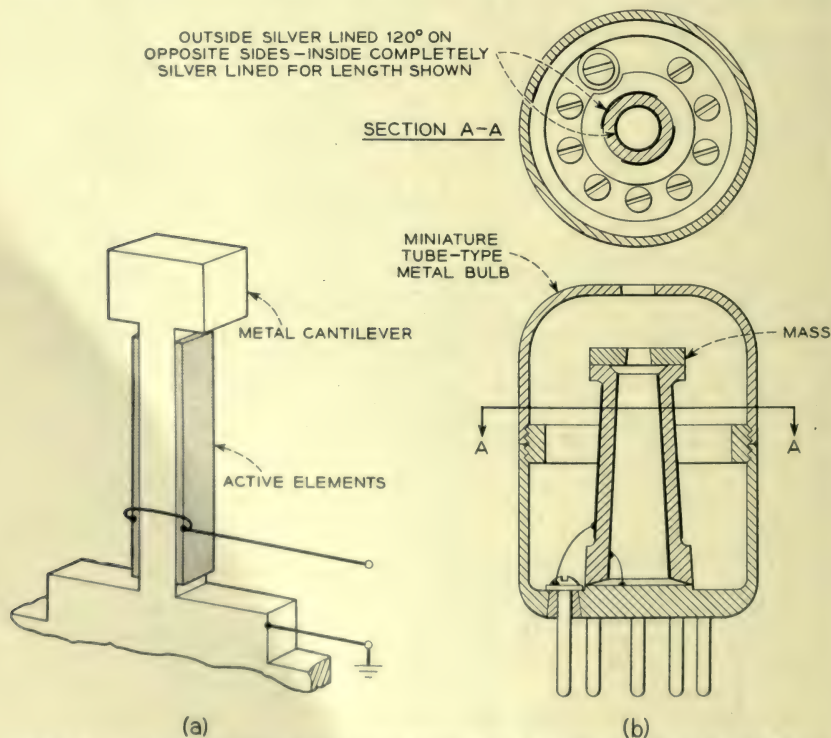


FIG. 6 — (a) Metal cantilever type element and (b) radial and axial miniature cantilever type accelerometer.

systems at amplitudes that are measurable with the help of a microscope. The peak acceleration in gravitational units for sinusoidal motion is given by:

$$g = .102 \text{ (c.p.s.)}^2 \times \text{amplitude (inches)}$$

At higher frequencies, this method of calibration becomes increasingly difficult due to the decrease in obtainable amplitudes. A calibration method for a wide frequency range is described in Reference 3. In general, the sensitivity of an accelerometer is a function of the active material, its size, and the weight M employed. The useful frequency range increases with decreasing sensitivity. For a given design, a compromise therefore has to be made between these quantities and the over-all permissible weight of the finished unit. To illustrate the approximate relations of weight, size and sensitivity, units have been constructed in the shape of miniature tubes weighing only 33 per cent more than their prototypes, with sensitivities in the order of 0.005V rms/g and a useful frequency range of 3,000 cycles.

Since these accelerometers are calibrated for rectilinear motion, the results obtained in measuring equipment vibrations or testing machines must be carefully interpreted. In many instances the disturbing forces impart a rocking motion to the units so that the position of the active elements in the housings will influence the acceleration magnitudes that are registered.

Associated Instrumentation

Since the voltage output of self generating accelerometers is small, electronic amplifiers have to be employed to bring the signal to desired levels. Fig. 7 illustrates a typical arrangement of the necessary equipment. A cathode ray oscilloscope is shown as the visual indicating means, although other recording instruments can be employed, depending on the nature of the disturbances to be measured and the type of record desired. The prime requirements of the equipment are:

(a) phase distortion must be low, so that the disturbance pattern is correctly presented.

(b) its transient and frequency response must be adequate for the frequencies to be recorded.

In the circuit shown in Fig. 7, the signal is fed into a cathode follower stage having a high input impedance in order to obtain good sensitivity and frequency response. The use of a cathode follower also offers more flexibility in the proper matching of the high impedance pick-ups to suitable low pass filters or to the following amplifying stage. The generated

signal is shunted by the capacity of the connecting cable. Since the voltage impressed on the grid of the first stage is approximately

$$V_2 = V_1 \frac{C_1}{C_1 + C_2}$$

it is desirable to have C_1 large compared to C_2 . In some types of investigations cable whip may result in the introduction of spurious voltages on the signal. These voltages are produced by capacitance changes and static charges on the cable dielectric. To minimize this effect, the cable may be shunted by a padding condenser at the expense of the voltage V_2 . The National Bureau of Standards recently reported the development of a low noise cable⁴ that is reported to overcome the shortcomings of the ordinary shielded cables. Cables with a sufficiently low noise figure have also become available commercially.

A timing and calibration voltage can be impressed across the low valued resistor R for comparison with the disturbance signal. If resonant frequency signals of the accelerometers are excited by the disturbances being investigated, or the recorded frequency range is to be limited, the signal is fed through low pass filters before being amplified. This prevents possible overloading of the amplifier by the unwanted frequencies, so that maximum gain can be realized for the desired signal.

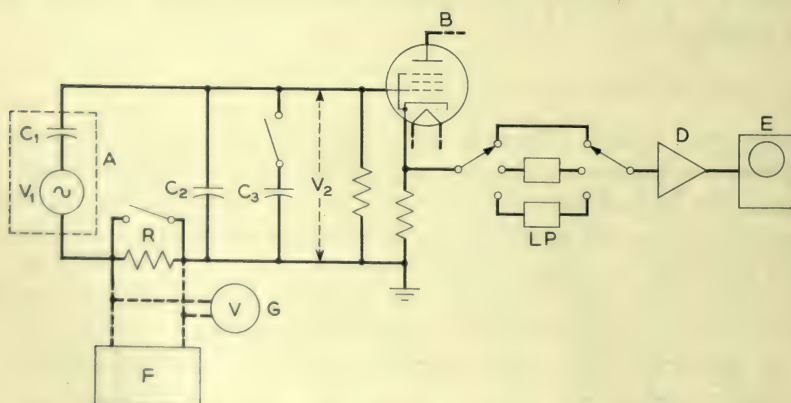


Fig. 7 — Schematic of accelerometer and associated recording circuit.

- | | |
|--|--|
| A — Equivalent circuit of accelerometer | D — Voltage amplifier |
| B — Cathode follower stage | E — Recording instrument |
| LP — Low pass filters | F — Audio oscillator |
| C_1 — Internal capacity of accelerometer | G — Voltmeter |
| C_2 — Cable capacity | R — Calibrating resistor |
| C_3 — Padding condenser | V_1 — Voltage source |
| | V_2 — Voltage impressed on first stage |

Particular attention must be paid to the design of the filters since these elements are apt to introduce excessive phase shift and oscillations. A variable frequency electronic filter has been found to be quite serviceable due to its flexibility in the choice of cut-off frequencies.

Additional information on shock and vibration instrumentation is given in Reference 5. Since the presentation of mechanical disturbances is, to some extent, limited by the transducers and their associated circuits, there is a trend towards a certain amount of standardization of these components so that results can be compared on an industry wide basis.

TUBE DESIGN PROBLEMS IMPOSED BY ENVIRONMENTAL CONDITIONS

It follows from the numerous acceleration measurements made at the installation points of equipment, that electron tubes, together with other components, have to withstand a large variety of conditions. Without over-simplifying the problems involved, it is perhaps permissible to divide these disturbances into two classes: ballistic shock, and transitory or sustained low g vibrations. Although, as a first approach, the effect of these disturbances on tube elements may be probed by mathematical analysis, the final design must be proven in by laboratory tests under controlled conditions. In discussing the influence of shock and vibration on tubes, their elements are frequently presented schematically as cantilevers, Fig. 8(a). While this assumption is a close approximation for some of the older tubes, such as the Western Electric No. 349B tube, Fig. 8(b), most tubes of later design, Fig. 8(c), do not lend themselves to such simple analysis due to their more complex structure. The response of elements to even simple shock pulses on tube envelopes are influenced by factors such as the clearances in micas, mica fits in the bulbs and tightness of mount assemblies.

It is obviously beyond the scope of this paper to analyze the destructive effects of shocks and vibrations on equipment. A few of the many excellent articles and publications on this subject are listed in Reference 6. It is equally impossible to present all of the many problems facing the electron tube engineer in designing tubes that will reliably serve their purpose under adverse conditions. Since tubes must be designed to withstand disturbances encountered in the field or be adequately protected, the following notes, highlighting some of these problems, will be of interest to equipment as well as tube engineers.

Influence of High g Shocks

In certain applications, tubes are required to withstand occasional high shocks such as those produced in military applications by explosions

of some kind. It is generally permissible that, during very severe shocks, electronic equipment is non-operative; therefore, the response of tubes during these disturbances may fall outside of assigned limits. However, since it is highly important that operations resume immediately, the tubes must show no permanent change nor have caused damage to the circuits as a result of temporary faulty operation. Although protective shock mounts are usually employed on equipments exposed to these conditions, damped equipment vibrations excited by the attenuated shock wave are superimposed on the pulse felt by the tube.

Brittle or stiff tube components are most susceptible to shock because of their inability to absorb the shock energy by elastic deformations. Failures falling into this category are:

(a) glass breakage, which may be brought about by impact due to excessive movement of the tube or adjacent components during the impact.

(b) metal to glass seal fractures, produced by shock loads on the tube leads and seals.

(c) heater or filament failures and opening of welds.

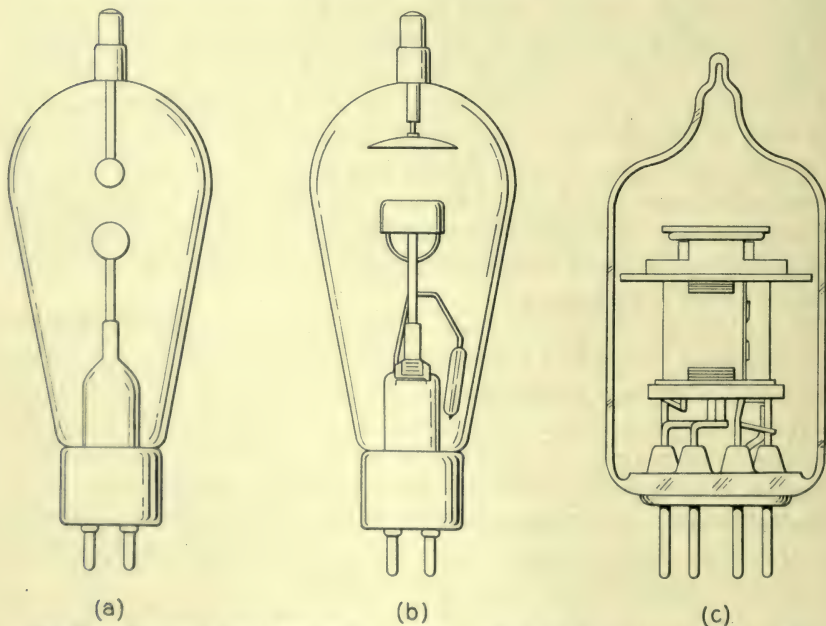


Fig. 8 — Electron tube structures: (a) Simplified anode and cathode structure of a Western Electric No. 349B tube, (b) Western Electric No. 349B tube, and (c) typical miniature tube.

(d) damage to mica serrations and enlarging of the mica holes which support and space tube elements. The deterioration of the mica is also known to liberate gas, which, in turn, results in a reduction of the vacuum.

Permanent damage is also caused by deformation of elements beyond their elastic limit. Bowing of grids and cathodes as a result of shock has been reported in some applications.

Although not always realized by design engineers, shocks and vibrations of surprising magnitudes may also occur in handling and shipping. If these conditions are not taken into consideration during the design stage, the over-all cost of the product may be adversely affected by the necessary protection that has to be built into the shipping container to assure safe arrival of the packaged article at its destination. While many factors enter into the proper design of shipping containers, such as moisture and corrosion protection, the selection of cushioning materials is perhaps the most important. Since it is desirable from a storage and shipping cost standpoint to keep the package bulk to a minimum, and protective packaging cannot always compensate for design weakness, adequate strength must be designed into the tube even though it will not be subjected to severe shocks once it is installed. A rather complete analysis of the dynamics of package cushioning is given in Reference 7. At present, military requirements specify that packaged tubes must safely withstand several three foot drops onto a hard surface.

Influence of Low g Disturbances

In contrast to the relatively infrequently occurring high peak shock and vibrations, we find that many equipments are often subjected to repetitive shocks and sustained vibrations at lower acceleration levels. These conditions are generally encountered in vehicle, ship and airplane applications. Although these shocks and steady state vibrations can be attenuated through the use of shock and vibration mounts, the effectiveness of these mounts may be reduced sharply by a change in disturbance frequencies from normal.

Tube failures resulting from these conditions are generally caused by fatigue of some tube elements. For instance, the continual hammering of micas against tube walls or chattering of cathodes and grids in the mica, may reduce the value of the micas as supporting and spacing elements, and since tubes are required to function under these conditions, the gradual degradation of the micas will bring about an increase in the tubes' microphonic output. Where microphonism is a factor, the useful

life of a tube is, therefore, terminated long before more apparent failures occur in the tube structure. Other points of weakness are heater and filament leads.

As may be deduced from the above, microphonism is a frequent cause of complaint when tubes are subjected to low but repetitive shocks or transient vibrations. To illustrate the influence of such disturbances on tube performances, the results of recent investigations of tube microphonism found in a certain equipment will be cited.

In this case, field reports indicated that certain tubes exhibited excessive microphonism in the equipment, although the tubes were found to be within limits when judged by standard factory tests. It was apparent, therefore, that these tests did not simulate actual conditions. Acceleration measurements made at the equipment base and at tube sockets revealed that the steep, short duration impact of a blow delivered to the outer case excited resonant vibrations of the chassis on which the tubes were mounted. The magnitudes of these vibrations were only in the order of 0.1g, but their lowest frequency (approximately 550 cps) was close to the mount resonances of some of the tubes. Further tests also showed that those tubes having pronounced response, i.e., low damping, to vibratory motion in this frequency range also proved to be microphonic in the equipment. (The vibration spectrum of one of the tubes is reproduced in Fig. 9.) It was found that the various modes of vibration of the tube mount produced the high peaks in the range between 500 and 1000 cps. Unfortunately, present factory tests do not include a complete evaluation of tube response over a wide enough frequency spectrum.

Observations made on several equipments indicate that structural changes in the chassis or re-positioning of tubes would, in some instances, reduce the effect of mechanical disturbances on tubes. An occasional source of trouble is introduced by equipment motor vibrations, especially after mechanical wear has increased play in the moving parts. The accelerations involved in these vibrations are usually very small, but if their frequencies coincide with structural resonances in the tubes, unsatisfactory operation of the equipment may result. The influence of such disturbances on tube operation is not always recognized by equipment designers.

Several programs are currently pursued by both military and commercial agencies to increase tube reliability. Since the necessary requisites that make a tube reliable depend on the type of service to be performed and environmental conditions, the requirements stressed in the various programs differ in many respects. Some of these requirements are still in a state of flux, as actual needs are as yet not clearly known.

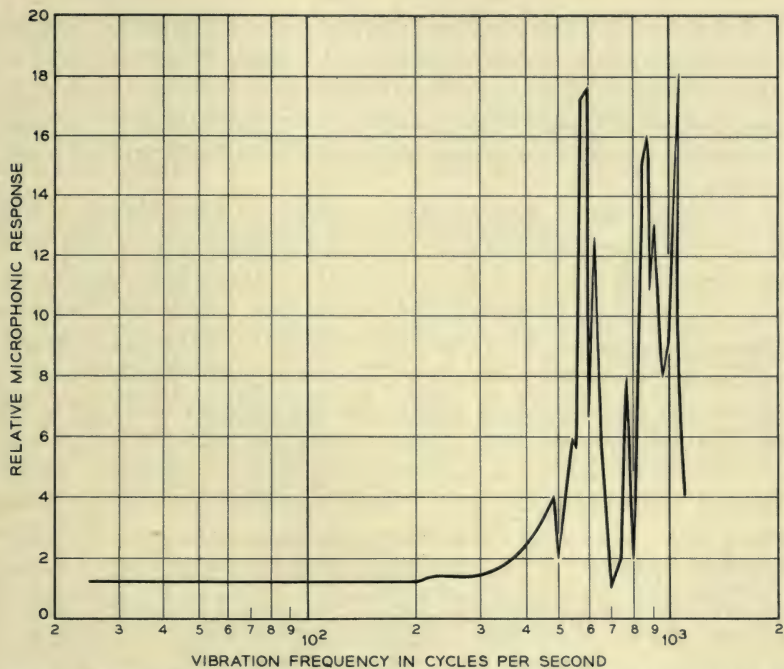


Fig. 9 — Frequency response of a tube that exhibited microphonism in equipment. The direction of vibration is perpendicular to the major diameter of grids and tube axis.

SHOCK AND VIBRATION REQUIREMENTS AND TEST EQUIPMENT

General requirements

Certain tests have been written into the MIL-E-1B⁹ and other specifications to control the shock and vibration characteristics of tubes, and to check on their behaviour under given mechanical excitation. Long usage has given these specifications some validity in that they control the quality of the product even though actual field conditions may not necessarily be simulated in the tests.

At present, one or more of the following three types of tests are called for on tube specifications:

(a) *Shock tests.* These are high acceleration tests to insure that tube structures will withstand occasional shocks of given maximum magnitudes. Since, in general, tubes are not required to function during high peak shocks, no operating voltages are applied to the tubes for this test. The post shock requirements are that the tube characteristics must not have drifted out of their limits. In many cases, the type of shock tester to be used is specified, because it is difficult to define the shock output of

many of these machines in terms of acceleration and frequency content, etc. The response of tube structures may be vastly influenced by shocks of the same nominal acceleration but differing acceleration wave forms. Shock tests are also performed on tubes during their initial development to obtain the degree of cushioning required for safe shipping and handling purposes.

(b) *Vibration tests.* These are low acceleration, fixed frequency tests. They are made on machines with sinusoidal displacement output. Short time duration 25-cycle — 2.5g, or 50-cycle — 10g tests are specified for tubes that are generally not exposed to constant vibrations. Usually no voltages are applied to the tubes under test since their prime purpose is to check for sound tube structures. Electrical post vibration performances are the criteria of tube quality. In this category also fall the long time duration vibration tests made on tubes intended for use in equipment that is known to be subjected to continual vibration, such as shipboard or mobile equipment. Present specifications call for 96-hour tests at 2.5g — 25 cycles. This, therefore, is a fatigue test which determines the capability of tube structures, under prolonged cyclic stresses.

(c) *Microphonic vibration test.* A 25-cycle — 2.5 g test performed with specified voltages on the tubes under test to investigate the influence of low acceleration vibration on the output of tubes. This test is specified on certain tubes, especially those used for audio applications. The permissible magnitude of spurious signals excited by vibration is limited on the respective tube specifications.

(d) *Tap tests.* These tests are performed for two purposes. One is for the detection of defects such as foreign particles between close spaced adjacent elements, damaged elements, or poor welds. Open and short testers of given sensitivity are used to indicate these defects. The second purpose for tap testing is the investigation of microphonic response of tubes to mechanical disturbances. For these tests the tubes are made to work in Class A amplifier circuits. Acoustic feedback may be used, so that the tube is not only subjected to the mechanical tap, but also to the sound of the tap excited microphonic signal which is reproduced through a loud speaker spaced at a given distance from the tube. Sustained microphonism can be produced by these means under certain conditions. The purpose of this type of test is to simulate conditions to which tubes are subjected in some equipments, especially those closely coupled to audio output components.

Equipment

The following is a brief description of the machines used for the performance of the above tests and their output characteristics.

MIL-E-1B Bump Tester

This is one of the earliest devices employed for shock testing of tubes under controlled conditions. In order to assure uniformity of results the MIL specifications give its physical dimensions. Fig. 10 illustrates the tester and its method of use. The magnitude of the shock and its duration is given by tube weight, shape of tube envelope contacted by the hammer, resilience of rubber pad on the hammer, and the angle (θ) through which the hammer is permitted to swing before striking the tube.

Although for the performance of the tests, only the angle (θ) is specified, the shock characteristics of this device have been investigated,¹⁰ so that shock magnitudes and durations for any tube may be computed from the parameters given above. A typical acceleration time curve is shown in Fig. 11. The simple bell shaped outline of the accelerogram is given by the non-linear spring characteristic of the rubber pad and the generally cylindrical shape of the tube envelope.

Shock Testing Mechanism per ASA-C39.3

This mechanism is also used to check and compare the resistance of tubes to mechanical shocks of predetermined magnitude and duration. In this device the sample to be tested is rigidly fastened on a platform. A steel leaf spring supported at both ends is attached underneath this platform. The test on the sample is performed by raising the platform to a certain height, allowing it to fall on a steel anvil, and then catching it on the rebound.

The shock magnitude is given by

$$G \text{ max} = \sqrt{\frac{2hk}{W}}$$

and its duration by

$$J = \pi \sqrt{\frac{W}{12Kg}}$$

where W = tableweight (lbs.)

h = height of fall (inches)

K = spring constant of leaf spring.

A number of leaf springs are available to produce the desired shock characteristics. A full discussion of this mechanism and its performance are covered in Reference 11. A slightly modified version of the tester, used by the Laboratories, together with a typical shock pulse, is shown in Figs. 12 and 13. It can be seen that the pulse is essentially of sinusoidal shape with higher frequencies superimposed on it. The fundamental frequency is produced by flexing of the leaf spring during its contact

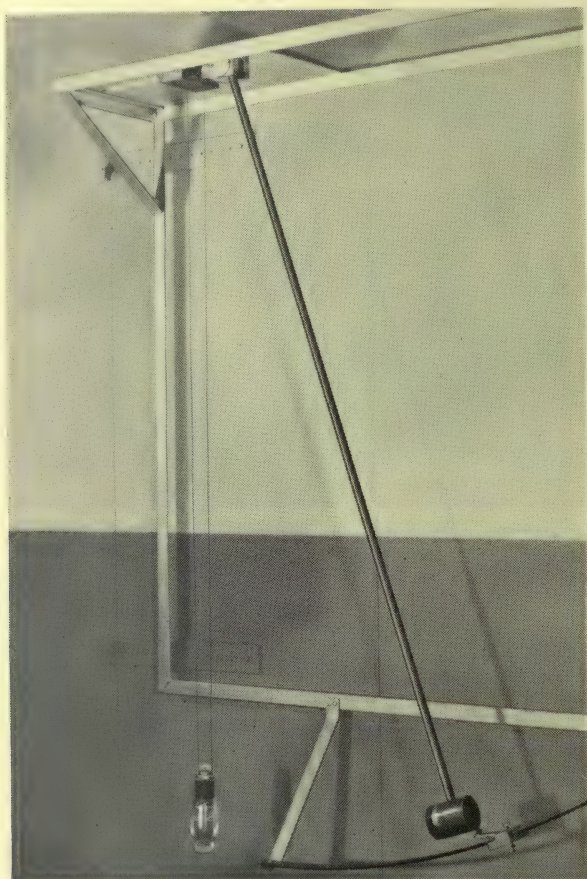


Fig. 10 — One of the early devices employed for shock testing of tubes under controlled conditions — the MIL-E-1B bump tester.

with the table anvil while the higher frequencies are due to table resonances which are excited by the metal-to-metal impact of spring and anvil.

High Impact Machine for Electronic Devices

This machine was originally intended to test the resistance of tubes to high peak-short duration shocks similar in nature to those encountered by equipment fastened to parts of ships that are likely to be exposed to direct explosion pressures.¹² The shock is produced by a steel hammer pendulum striking a movable steel table on which the tube under test is mounted. The shock magnitude is given by the angle through which the

pendulum is allowed to swing under the action of gravity before the hammer strikes the table. The forward motion of the table produced by the impact is arrested by two shock absorbers.

The impact of the hammer on the anvil of the table produces an abrupt velocity change of the table and excites table resonances, both horizontally and vertically. Because of the structure of the table a very complex acceleration wave form results. In general the accelerations for a given hammer swing vary over the table surface. It is for this reason that, in testing procedures, the position of the tubes on the table and their method of clamping are well defined; and since the acceleration wave shape is the sum of many vibratory frequencies rather than a single shock pulse, the severity of the test is expressed by the angle of hammer swing instead of a shock magnitude and duration. Although some uniformity in performance is attained by rigid standardization of the structure of the machines and the above mentioned positioning of the tubes, minute differences in these parameters may produce sufficient variations in shock output to reflect on test results. Acceleration-time traces of shocks measured in the shock direction by an accelerometer fastened near the anvil of the table and a second accelerometer clamped to the center of the table, are shown in Figs. 14 and 15. The records were simultaneously taken through 10,000-cycle low pass filters. Even though the recording of accelerations containing high frequencies of large amplitudes is, to some extent, a function of the measuring equipment, it is seen that significant differences in output exist between the two points of measurements.

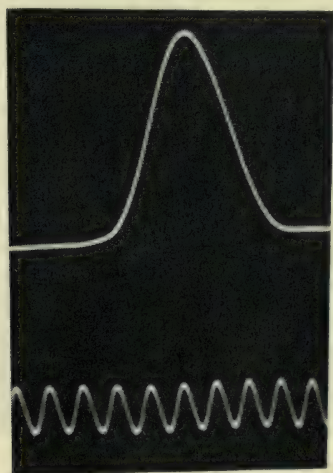


Fig. 11 — Acceleration-time pulse produced by MIL-E-1B bump tester.

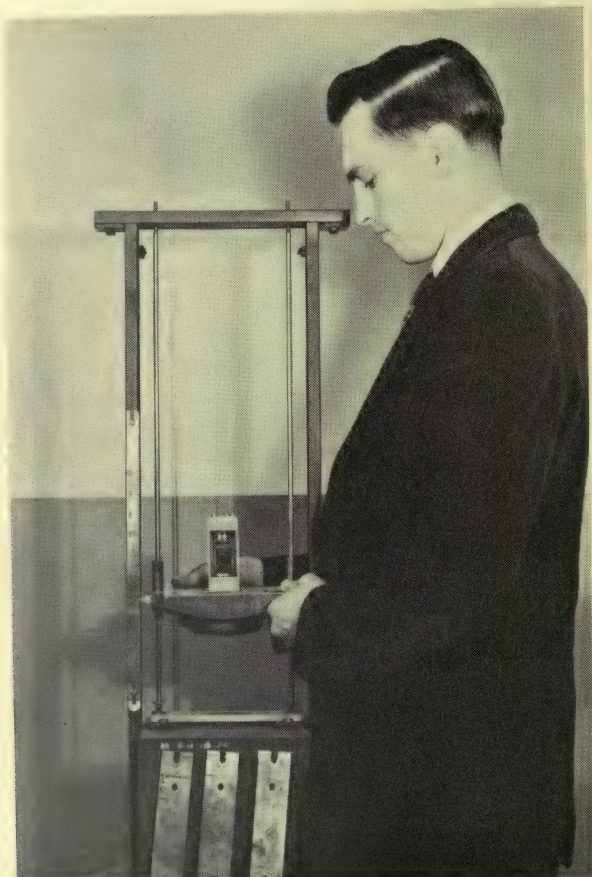


Fig. 12 — Shock testing mechanism per ASA-C39.3.

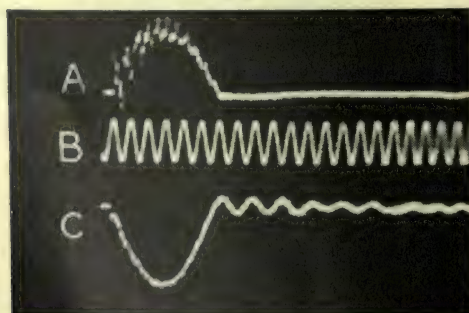


Fig. 13 — Typical shock pulse obtained with the tester shown in Fig. 12 (A) acceleration pulse, (B) timing trace, and (C) stress in leaf spring.

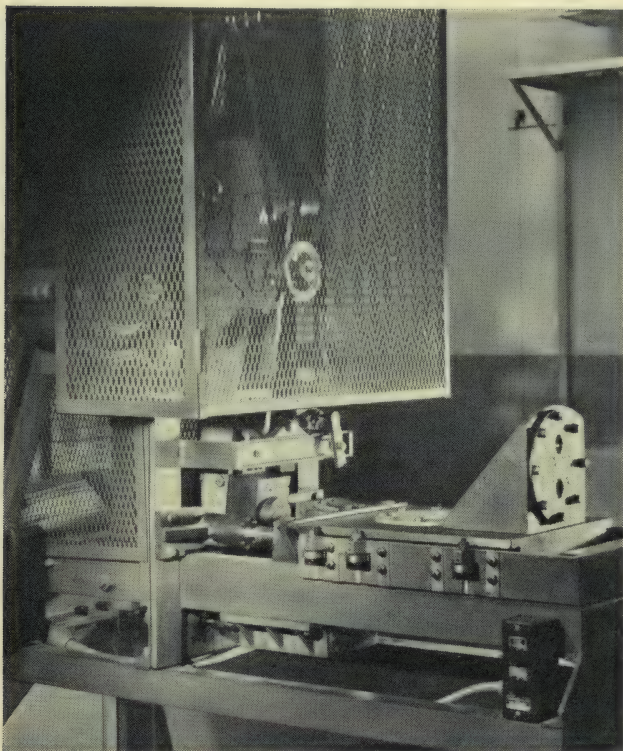


Fig. 14 — High-impact machine for testing electronic devices.

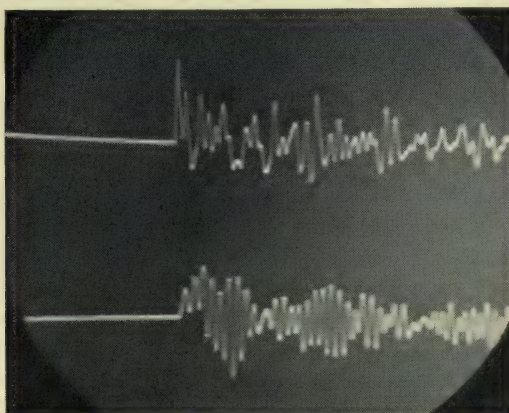


Fig. 15 — Acceleration-time pulses produced by machine shown in Fig. 14. Upper trace recorded near anvil, lower trace recorded in center of table.

This machine can also be used to produce lower G shocks of longer duration and simpler wave form by interposing a resilient rubber pad between the anvil and hammer. The resulting shock resembles the bell shaped output of the MIL-E-1B bump tester. Its magnitude and duration are given by the hammer swing, resilient characteristics of the pad, and the table and specimen weight. The capacity of the machine permits shock testing of heavier tubes (such as the larger magnetrons) by this method. Attainable shock levels are sufficiently high to cover the requirements placed on these tubes by conditions encountered during transit.

The L.A.B. Package Tester

Of particular interest to the packaging engineer, this machine is intended to duplicate the destructive vibrations and shocks experienced by a product during transit. It consists of a horizontal table that can be made to vibrate with a circular motion in the vertical plane. Adjustments permit variations of this motion to simulate freight car and motor truck movements. Tests performed on this machine, therefore, check on the mechanical strength of the outer shipping container as well as on the adequacy of cushioning materials employed to protect the product. The services are also considering this machine to test equipment designed for use in vehicles. Tests are now in progress by the Signal Corps to determine proper parameters for this application.⁸

Vibration Machines

A number of vibration machines, made by various manufacturers, are employed for vibration testing of tubes. Due to the high hash output of most mechanically driven machines, which contain gears and linkages, great care must be taken in the selection of these machines. Certain tests, especially the microphonic tests on receiving tubes, require machines with good sinusoidal output, in order to obtain comparable results. It is for this reason that the leaf spring vibration machine (Fig. 16), developed several years ago by Bell Telephone Laboratories, has been recommended as a standard for performing vibration tests. This is a fixed frequency, 25-cycle — 2.5 g machine which, due to its construction, has a relatively clean output as shown in Fig. 17.

Several types of electronically or motor-generator driven vibration machines are on the market. These are variable frequency and variable amplitude machines especially useful for determining resonance frequencies of structures and for performing cycling vibration tests. Accessory equipment has been made available lately to conduct these tests on an automatic basis at either constant acceleration or constant am-

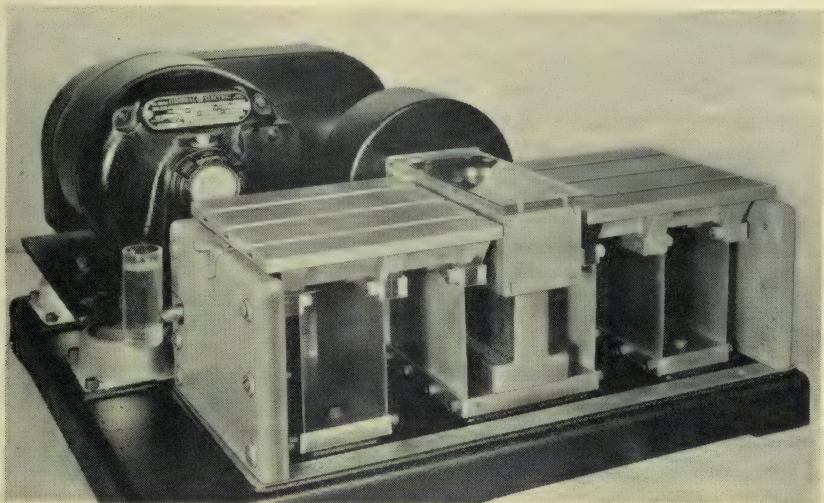


Fig. 16 — Leaf spring vibration machine developed several years ago by Bell Telephone Laboratories.

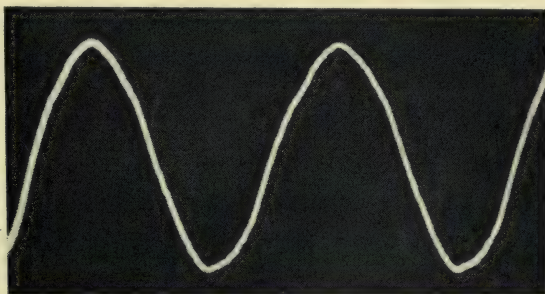


Fig. 17 — Acceleration output of the leaf spring vibration machine.

plitude over a given frequency range. Within the limitations of these machines and their power sources, it is also possible to subject specimens to complex disturbances. For instance, a projected use is the reproduction of recorded equipment vibrations so that the response of tubes to these conditions can be studied in the laboratory.

Tube Tappers

The physical construction of tube tappers used in the electron tube industry varies from the simple manually operated cork mallets to the more complicated automatic tappers. Since it has long been recognized that the reproducibility of manual tapping is rather poor, efforts have been made to replace these devices by automatic tappers or other

methods of checking microphonism in tubes. The industry, in collaboration with the Services, is now engaged in standardizing on automatic tappers. The prime requisites for such tappers are: (a) their shock output must be reproducible and must fall within definable limits, (b) their operation cycle must not adversely affect the time required for performing tap tests, and (c) the results obtained must have some relation to the environmental requirements for the tubes. This last condition is, perhaps, the most difficult to attain considering the diverse conditions encountered by tubes in various equipments.

Of the many tappers devised and employed by the industry, the use of only two are at present approved in the MIL-E-1B specifications. One is a manually operated cork mallet. This mallet is still retained in spite of its shortcoming, for lack of better devices. The second is the General Electric Automatic Tapper, specified for checking microphonism of some of the reliable tubes. This is a motor driven tapper which subjects the tube under test to damped low "g" vibrations at the rate of 2 taps per second. The tapper and associated circuits are fully described in the MIL specifications.

Space does not permit the discussion of other proposed tappers or test methods. Two rather interesting papers are listed in References 13 and 14, which illustrate the problems involved in the development of suitable devices for checking microphonism in tubes and describe some of the various methods of approach.

SUMMARY

The rapid growth of electronic equipment development during the late war has continued with the ever increasing new applications to both military and civilian purposes. It is realized that this growth has placed more stringent requirements on the mechanical characteristics of electron tubes. In order to define intelligently tube requirements, the nature of the disturbances to which tubes may be subjected must be known. It was pointed out that merely applying equipment requirements to the tubes, is not very realistic, since the shock and vibration patterns may be vastly modified by the equipment structures.

With the help of the recent development of light weight accelerometers, it is now possible to investigate the disturbances at the tube sockets. Both the Industry and the Services have begun to utilize these instruments to collect this information. The benefits derived from this work will enable the equipment and tube designers to formulate more accurate requirements and to devise test gear that will simulate more closely field conditions to check on tube quality.

Although, as a result of the trend to miniaturization, the strength of electron tubes has been increased, due to the smaller size and mass of elements employed, much has still to be done to increase tube resistance to ballistic shock. This would result in simplification of shipping containers and reduce the need for protection by shock absorbers in equipment. Equally important is the effort now made to reduce the microphonic response of the tubes. Here, too, the smaller size of the late tubes is of advantage because element resonant frequencies are increased. And since the higher frequency components of disturbances are largely attenuated by equipment structures, the tube responses have been lessened.

With a better understanding of the problems involved in tube protection, it is also quite possible that further improvements can be effected in some instances by structural changes of equipment members, or re-orientation of tubes in the equipment to reduce the effects of shock and vibrations on tubes.

BIBLIOGRAPHY

1. Shock and Vibration Instrumentation, NRL report #0-2645.
2. Mechanical vibrations, Den Hartog, McGraw-Hill Book Co.
3. The reciprocity calibration of piezo-electric accelerometers, M. Harrison, A. O. Sykes, and P. G. Marcotte, The David Taylor Model Basin, Report 811.
4. A Noise Free Instrument Cable, National Bureau of Standards, Technical News Bulletin, **36**, No. 3.
5. Research and Development Toward Construction of Accelerometers, Gulton Mfg. Co. report. New Techniques for Measuring Forces and Wear, W. P. Mason, B.S.T.J., **31**, May, 1952. Proceedings of symposium on Barium Titanate Accelerometers, National Bureau of Standards, Report No. 2654, Aug. 1953.
6. Bureau of Ships Publications, Nav. Ships 900,036, and Nav. Ships 900,052. The David W. Taylor Model Basin, Report 481. Vibration and Shock Isolation, C. E. Crede, John Wiley & Sons. Response of Damped Elastic Systems to Transient Disturbances, R. D. Mindlin, F. W. Stubner, H. L. Cooper, Experimental Stress Analysis, **5**, No. 2. Measured Aircraft Vibrations as a Guide to Laboratory Testing, D. C. Kennard, Jr., A. F. Technical Report No. 6429.
7. Dynamics of Package Cushioning, R. D. Mindlin, B.S.T.J., **24**, July — Oct., 1945.
8. Evaluation of Electron Tube Quality on the Tethered Weapons Carrier, Signal Corps Engineering Laboratories, Army Project No. 0302A.
9. MIL-E-1B, Military Specifications for Electron Tubes.
10. Characteristics of the JAN Bump Tester, R. D. Mindlin, F. W. Stubner and H. A. Lefkowitz, MM-45-2920-4.
11. Shock Testing Mechanism for Electrical Indicating Instruments, ASA-C 39.3
12. Mechanical Shock Characteristics of the High Impact Machines for Electronic Devices, I. Vigness, R. C. Novak and E. W. Kammer, NRL Report No. 0-2838.
13. Proposal for Revised Noise and Microphonic Evaluation for MIL-E-1B Specifications, NML Report, Dated July 30, 1951.
14. A Study of Microphonism, Philco Research Report No. 183.

Arcing of Electrical Contacts in Telephone Switching Circuits

Part I—Theory of the Initiation of the Short Arc

By M. M. ATALLA

(Manuscript received April 1, 1953)

This is a presentation of a theory for the mechanism of the initiation of the short arc commonly observed on the closure and opening of electrical contacts. The theory is based on the experimental evidence that an established arc is generally preceded by a period of local high frequency discharges at the contacts. During this period the circuit current builds up. If and when this current reaches the arc initiation current of the contact a steady arc is established. It is shown that this initiation period of the arc is directly determinable from the circuit conditions and the contact condition. This mechanism furnishes rather simple explanations to some complex phenomena commonly observed before the establishment of the arc. The mechanism of the initiation of an individual discharge, however, still remains uncertain.

INTRODUCTION

In the course of study of arcing phenomena between electrical contacts, it has been long established that a condition for sustaining the short arc is to maintain a current through the arc greater than a minimum value called the minimum arcing current. This current is generally a characteristic of the contact material and is appreciably affected by surface contaminations. For clean metals the minimum arcing current is usually equal to a few tenths of an ampere. Before establishing the arc, therefore, there must exist a certain mechanism which accounts for a rapid current build-up from zero to a value as high as the minimum arcing current. For an inductive circuit, a higher inductance should result in a longer period of current build-up. With a sufficiently high circuit inductance this initiation period may be made long enough to be directly observed and to allow an examination of the mechanism involved. Such experiments have been made and observations have indicated that the initiation period consisted of a succession of rapid dis-

charges at the contacts from the open circuit voltage to a lower voltage. During this process the current in the circuit built up in a discontinuous fashion and the steady arc was established only when the circuit current reached the minimum arcing current of the contact.

In this paper is presented our study of this initiation period. The following are the main objectives: (1) to establish the analytic relations governing the performance of a few simple contact circuits during the arc initiation period; (2) to check the analysis by direct measurements; (3) to apply the theory to explain a few arcing phenomena and empirical relations previously reported; and (4) to shed some light on the nature of the rapid local discharges at the contacts and the characteristics of the influential circuitry at the immediate neighborhood of the contact.

NOTATION

C	Main circuit capacitance
I	Circuit current
I_i	Arc initiation current
I_m	Arc termination current or minimum arcing current
L	Circuit inductance
$(L)_{Limit}$	Limiting or maximum inductance above which a steady arc cannot be established
R	Circuit resistance
V	Voltage
V_0	Initial voltage
V_{cr}	Main condenser terminal voltage
c	Local capacitance at the contacts
ℓ	Local inductance at the contacts
n	Number of discharges at the contacts
r	Local resistance at the contacts
t	Time
v	Voltage across a steady short arc
$\bar{v}$	Voltage across the contacts at the termination of a single local discharge
z	Local impedance at the contacts $\left[\frac{\ell}{c} \right]^{1/2}$
α	Ratio of capacitances c/C
ω	Angular frequency $(Lc)^{-1/2}$

ANALYSIS

In this section a few simple contact circuits are considered. In each case relations are derived for the current and voltage changes in the

circuit during the period of rapid discharges at the contacts preceding the steady arc. The analysis is based on the following simplified model of the mechanism involved: (1) The first local discharge at the contact takes place when the proper separation corresponding to the initial voltage V_0 is reached; (2) this discharge time is assumed to be short and negligible in comparison to the following charging time; (3) the local capacitances at the contact recharge from the main circuit until the same initial voltage V_0 is reached when a second discharge takes place; (4) this process repeats until a steady arc is established provided that the circuit is capable of building up enough current, — otherwise, the local discharges will continue and finally stop when the main circuit becomes incapable of charging the local contact capacitances to V_0 ; and (5) all the local discharges at the contact are terminated at a constant voltage $\bar{v}$ for any one set of circuit conditions. The nature of $\bar{v}$ is left to be determined and physically understood from our measurements.

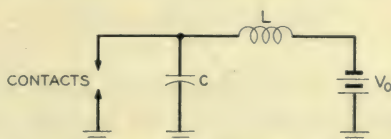


Fig. 1 — Typical battery-inductance-contacts circuit.

Battery V_0 , L and Contacts, Fig. 1

Following the first discharge at the contact from V_0 to $\bar{v}$ the local contact capacitances will recharge with a current

$$I = (V_0 - \bar{v}) \left(\frac{C}{L} \right)^{1/2} \sin \omega t,$$

where $\omega = (Lc)^{-1/2}$.

The voltage at the contact will reach V_0 at $t_1 = \pi/2\omega$ and the corresponding current is

$$I_1 = (V_0 - \bar{v}) \left(\frac{C}{L} \right)^{1/2}.$$

A second discharge will then take place and recharging will proceed with the new boundary conditions: at $t = 0$, the contact voltage is $\bar{v}$ and the circuit current is I_1 . By following this procedure, a general expression for the n th charging process is obtained:

$$I(n) = (V_0 - \bar{v}) \left(\frac{C}{L} \right)^{1/2} (n)^{1/2} \quad (\text{a})$$

and

$$\omega \cdot t(n) = \cot^{-1} (n - 1)^{1/2} \quad (1)^*$$

The relation between the number of discharges n and time t is given by

$$t = \frac{1}{\omega} \sum_{n=1}^{\infty} \cot^{-1} (n - 1)^{1/2}.$$

Only an empirical expression for this summation was obtained with a maximum error less than 10 per cent:

$$t = \frac{\pi}{2\omega} (n)^{1/2}.$$

The current build-up in the circuit is, therefore, expressed as a function of time by the following relation:

$$I(t) = \frac{2}{\pi L} (V_0 - \bar{v})t \quad (2)$$

In other words, the current is independent of the contact capacitances and increases linearly with time at a rate inversely proportional to the circuit inductance.† As soon as the circuit current reaches the arc initiation current I_i , a steady arc is established. The initiation time of the steady arc is therefore given by:

$$t_i = \frac{\pi}{2} \cdot \frac{I_i L}{V_0 - \bar{v}} \quad (3)$$

For $I_i = 0.5$ ampere, $L = 10^{-6}$ henry and $V_0 - \bar{v} = 30$ volts, $t_i = 2.6 \times 10^{-8}$ second.

C, L and Contact, Fig. 2

If the battery is replaced by a condenser C where the ratio $\alpha = c/C$ is much less than 1.0 an analysis similar to the above can be made. In

* The assumption that all discharges will take place from the same voltage V_0 is only true if: the motion of the contact is negligibly small, the discharges do not change the contact geometry to the extent of materially changing the contact separation, and if the effect of the residual ions is negligible.

† This is only true if the circuit resistance is zero. For a finite circuit resistance R , and $\frac{R}{2} \left(\frac{c}{L} \right)^{1/2}$ much less than 1.0, it can be shown that the current approaches the asymptotic value $\frac{V_0 - \bar{v}}{2R}$.

‡ It is shown later by measurement, that the initiation current I_i is essentially the same as the arc terminating current I_m .

this case, however, in setting the boundary conditions one must consider the drop in voltage across the main condenser during the previous charging processes. The following are the resulting expressions for the circuit current, main condenser voltage and the charging time, all as functions of n :

$$I(n) = \frac{V_0 - \bar{v}}{L\omega} \cdot \left[n(1 - \alpha(n - 1 + \alpha n)) \right]^{1/2} \quad (4a)$$

$$V(n) = V_0 - \alpha n(V_0 - \bar{v}) \quad (4b)$$

$$\omega \cdot t(n) = \sin^{-1} \alpha \left[\frac{n}{1 - \alpha(n - 1)} \right]^{1/2} + \tan^{-1} \left[\frac{1 - \alpha(n - 1)}{(1 + \alpha)(n - 1)} \right]^{1/2} \quad (4c)$$

Only an empirical expression for the summation

$$\sum_{n=1}^{n=1} t(n)$$

was obtained with an error less than 10 per cent

$$\sum_{n=1}^{n=n} \omega \cdot t(n) = \frac{\pi}{2} (n)^{1/2} (1 + \alpha n).$$

The current relation indicates that the current increases from zero at $n = 0$ to a maximum current

$$I_{\max} = \frac{V_0 - \bar{v}}{2} \left(\frac{C}{L} \right)^{1/2}$$

at $\alpha n = 1/2$ then drops back to zero at $n\alpha = 1.0$ when the discharges are terminated. The total discharge time is approximately $\pi(LC)^{1/2}$ and the terminal voltage on the main condenser is $V_{CT} = \bar{v}$. If during the process of current build-up the current reaches a value equal to the arc initiation current a steady arc is established. It is evident that a steady arc cannot be established if the maximum current attainable during the discharges is less than the arc initiation current. This leads to the concepts of a limiting inductance and limiting voltage in a circuit that can allow the establishment of a steady arc.¹ The limiting inductance is

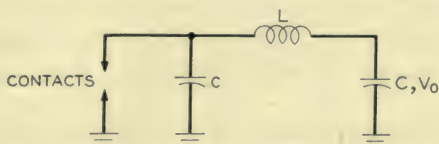


Fig. 2 — Typical condenser-inductance-contacts circuit.

¹ L. H. Germer, Arcing at Electrical Contacts on Closure, Part I, J. Appl. Phys. **22**, p. 955, 1951.

defined as the largest inductance for a given circuit above which a steady arc cannot be obtained and the limiting voltage is defined as the lowest voltage for a given circuit below which a steady arc cannot be obtained:

$$(L)_{\text{Limit}} = \frac{C}{4} \cdot \left(\frac{V_0 - \bar{v}}{I_i} \right)^2 \quad (5a)$$

$$(V_0 - \bar{v})_{\text{Limit}} = 2I_i \left(\frac{L}{C} \right)^{1/2} \quad (5b)$$

The assumption made that all charging processes are much longer in time than the discharging times, imposes a limitation on the applicability of the above relations. Equation 4c can show that the minimum charging time is $(2/\omega)\alpha^{1/2}$. The accuracy of the above relations is better the larger this time is compared to the discharge time of the local contact capacitance. Assuming distributed characteristics for the local and relatively small contact circuitry* the discharge time is $2(\ell c)^{1/2}$. The limitation involved, therefore, is that $(2/\omega)\alpha^{1/2}$ must be greater than $2(\ell c)^{1/2}$ or $L/C > \ell/c$. In other words, the impedance of the main circuit must be greater than the impedance of the local circuit at the contact.

MEASUREMENTS

All the measurements presented were obtained from an inductive type circuit, *C-L-Contact*. The main circuit, Fig. 3, consisted of a condenser *C* in series with a honeycomb inductance which is connected by a short lead, about 2 cms, to a pair of clean palladium contacts operating in laboratory air. The contacts were mounted on a cantilever bar arrangement, described by Pearson², which allows fine adjustments of the separation between the contacts as well as slow motion of the contacts to avoid physical closure before the end of a transient.

Three sets of measurements were made.

(1) Contacts voltage measurements: the transients obtained usually needed some correction to compensate for the effects of the measuring oscilloscope circuit. None of these measurements are presented in this paper. It may be mentioned, however, that they indicated the existence of rapid discharges at the contacts preceding the establishment of the steady arc.

(2) Circuit current measurements, Fig. 3 (a): these were made by

* In a later section of this paper, measurements were shown to indicate the plausibility of this assumption.

² G. L. Pearson, Phys. Rev. **56**, p. 471, 1939.

measuring the voltage change across a 10-ohm non-inductive resistor inserted between the main circuit condenser and ground.

(3) Main condenser voltage measurements, Fig. 3(b): the oscilloscope plates were shunted by an 1100×10^{-12} farad capacitor and the combination was used as the main circuit condenser. The above sets of measurements 1 and 2 furnished the data necessary for a quantitative establishment of the theory. It is evident that the measuring scope circuits did not interfere with the normal behavior of the contact circuit.

Development of Circuit Current During Arc Initiation Period

The circuit in Fig. 3(a) was used. The contact separation was gradually decreased until the discharges started and the resulting current transients were recorded. The circuit parameters were chosen such that, according to Equation 3 of our analytical results, the initiation time was of the order of microseconds. Fig. 4(A) shows a typical current-time transient where a steady arc was established. The circuit had $L = 1100 \times 10^{-6}$ henry and $C = 10^{-6}$ farad. The current started from zero and increased during the multiple discharge period until point 1 where the current was 0.2 ampere and a sustained arc was established. The current then increased to a maximum of 1.7 amperes then dropped. At point 3 the arc stopped when the current was 0.24 ampere. In Fig. 4(B) the first

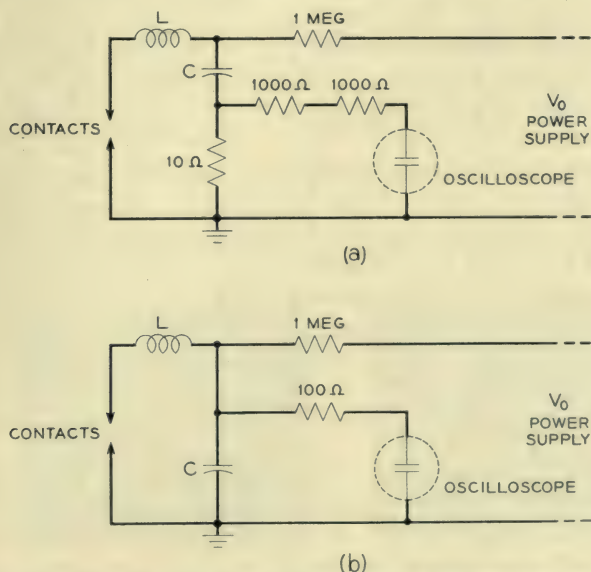


Fig. 3 — (a) Contacts circuit and circuit current measuring circuit. (b) Contacts circuit and main condenser voltage measuring circuit.

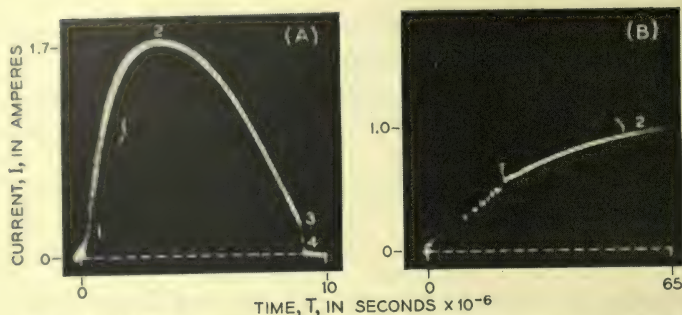


Fig. 4 — Circuit current transients with steady arc established.

portion of a similar transient is shown: $L = 5100 \times 10^{-6}$ henry and $C = 10^{-6}$ farad. The current build-up during the multiple discharge period is shown as an interrupted line 0-1. The steady arc was established at 0.57 ampere.

From a group of similar transients, pairs of measurements were made of the initiating current and terminating current of the steady arc. The results are given in Table I. It is concluded from the results that *the initiating current and terminating current of an arc are the same.*

The high-frequency discharges at the contact do not necessarily have to be followed by a sustained arc. *A sustained arc is only obtained if the maximum current established during the local discharges at the contact is equal to or greater than the arc initiation current.* Fig. 5 (A) shows a current-time transient without an initiation of the steady arc: $V_0 = 400$ volts, $L = 5100 \times 10^{-6}$ henry and $C = 1100 \times 10^{-12}$ farad. The maximum current reached was only 0.13 ampere which was not sufficient for initiating a steady arc. It is of interest to notice that the oscillations superimposed on the zero current line following the transient can allow a calculation of the local capacitances at the contact. Such a calculation gave $c = 7.8 \times 10^{-12}$ farad. Fig. 5 (B) shows a similar transient for the same circuit with $V_0 = 500$ volts. Following the first current build-up and drop, 1-2-3, the main condenser had a residual negative voltage high enough to produce a reversed current build-up and drop, 3-4-5.

Voltage Drop in Main Condenser During Arc Initiation Period

During the period of rapid discharges the current build-up in the circuit is accompanied by a voltage drop at the main condenser. This drop may correspond to one of two phenomena at the contact: (1) multiple discharges leading to a steady arc, and (2) multiple discharges without steady arc, followed by an open circuit. Fig. 6 and 7 are re-

TABLE I — ARC INITIATION AND ARC TERMINATION CURRENTS
PALLADIUM CONTACTS IN AIR*

Arc Initiation Current I_i : Amps.	0.12	0.21	0.21	0.15	0.25	0.19	0.23	0.40	0.65	0.61	0.57	0.52	0.45
Arc Termination Current I_m : Amps.	0.13	0.16	0.17	0.20	0.22	0.24	0.24	0.42	0.50	0.52	0.53	0.58	0.59

* Both numbers given in one column were obtained from the same transient.

spective records of the voltage change at the main condenser. Fig. 6 (A) corresponds to the case where the multiple discharge period was short and lead to a steady arc. During this arc the main condenser voltage dropped from point 1 to point 2 when the arc was arrested. This was followed by striking a second arc in the opposite direction which lasted until point 3. Line 3-4 represented the recharging of the main condenser from the power supply circuit. Superimposed on the same figure is the trace 1-2-3-5 corresponding to a closure at the contact instead of an open circuit. While line 1-2-3 shows two consecutive arcs, it is generally possible to obtain any number of such arcs. An even number of arcs will result in a *positive* residual voltage at the main condenser, Fig. 6 (A) while an odd number of arcs will result in a *negative* residual voltage at the main condenser, Fig. 6 (B).

Figs. 7 (A) and 7 (B) correspond to the case when the multiple discharges did not lead to a steady arc due to insufficient current build-up. The multiple discharges caused a voltage drop 1-2 across the main condenser followed by an open circuit and recharging, 2-3.

Discharge of the Local Circuitry at the Contact

Fig. 8 (a) represents a plausible representation of the local circuitry at the contact. When the conditions between the contacts are favorable

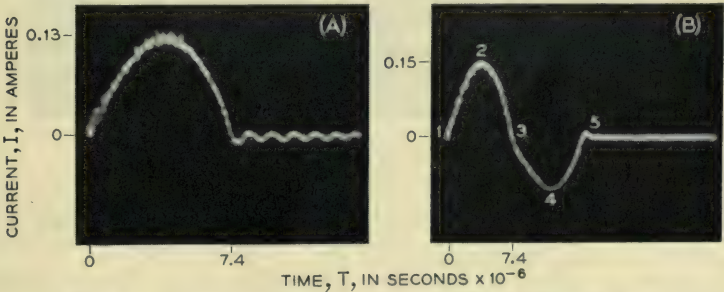


Fig. 5 — Circuit current transients without a steady arc.

for the initiation of the arc, a small local capacitance c' will furnish the necessary charge, through a small impedance z' , according to whatever mechanism that may be involved in the initiation process.* The connection from the contact to the main circuit is represented by a short transmission line with the distributed characteristics, r , ℓ and c . The drop at the contact from the initial voltage V_0 to the arc voltage v will cause a current surge $(V_0 - v)(c/\ell)^{1/2}$, if r is neglected, for a period $2(\ell c)^{1/2}$ corresponding to the time required by the pulse to travel to the end of the line and return to the contact. At this time the arc is extinguished by the reflected pulse and the final voltage at the contact is $-(V_0 - 2v)$. Figs. 8(b) and 8(c) are diagrammatic representations of the process. For a purely inductive line, therefore, the contact voltage $\bar{v}$ following one discharge is $-(V_0 - 2v)$. For a dissipative line, however, $\bar{v}$ is algebraically greater. Equation 4 of Germer and Haworth³ was derived to give the voltage following an arc for a similar circuit with lumped characteristics. For $r/(\ell/c)^{1/2}$ less than 1.0 this equation can be

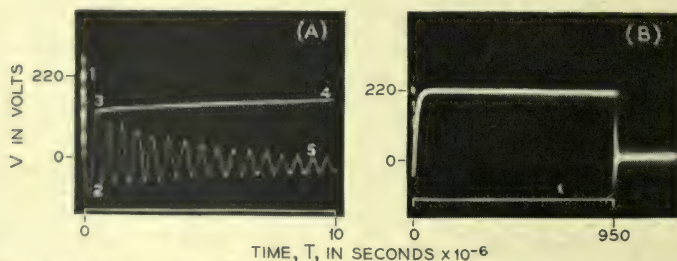


Fig. 6 — Voltage transients across main condenser with steady arc established.

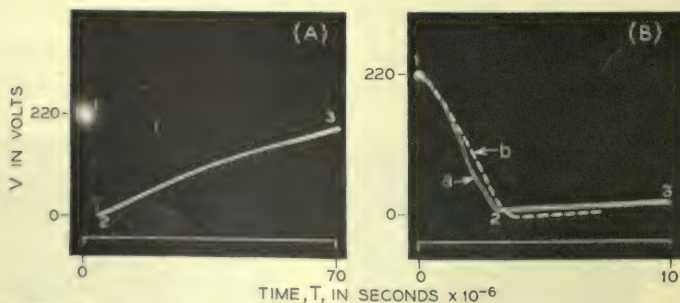


Fig. 7 — Voltage transients across main condenser without steady arc.

* This mechanism is not clearly understood at the present time. It is the opinion of the writer, however, that the initiation time is only a few times the transit time of the metal ion under the prevailing conditions.

³ L. H. Germer and F. E. Haworth, Erosion of Electrical Contacts on Make, Appl. Phys. 20, p. 1085, 1948.

approximated by:

$$\bar{v} = 2v - V_0 + \frac{\pi}{2} \cdot \frac{r}{z} (V_0 - v) \quad (6)$$

where $z = (\ell/c)^{1/2}$.

Our measurements were then applied to demonstrate the plausibility of the above description of the local circuitry at the contact. This was done in the following fashion. $\bar{v}$ was obtained by three methods. (1) Measurements were made of the multiple discharges preceding a steady arc, from records similar to Fig. 4(B), and $\bar{v}$ was calculated from Equation 3, (2) measurements were made of the maximum current attainable during the multiple discharge period, from records similar to Fig. 5(A), and $\bar{v}$ was calculated from the expression

$$I_{\max} = \frac{1}{2} (V_0 - \bar{v}) \left(\frac{C'}{L} \right)^{1/2},$$

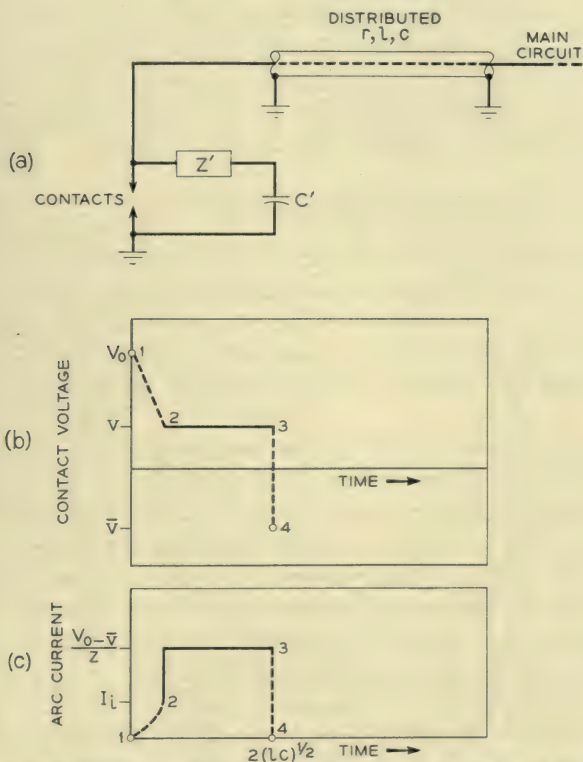


Fig. 8 — (a) Representation of local circuitry at contacts. (b) Voltage transient at contacts during one discharge of local circuitry. (c) Current transient at contacts during one discharge of local circuitry.

TABLE II — RATIO r/z FOR LOCAL CIRCUITRY AT CONTACTS

Initial Voltage V_0	100	150	200	230	250	300	360	400	460
$\frac{r}{z}$	0.56	0.65	0.55	0.41	0.55	0.30	0.40	0.40	0.48

and (3) measurements were made of the terminal voltage across the main condenser at the end of the multiple discharges, from records similar to Fig. 7, which, according to our analysis, is equal to $\bar{v}$. For these values of $\bar{v}$ Equation 6 was used to compute r/z . The results are given in Table II. Each value of r/z in this table is the average of 3 to 6 values obtained by the different methods described above. In all cases r/z was between 0.4 and 0.7. This indicates that the local contact circuitry is oscillatory. For our circuit, c was measured at 7.8×10^{-12} farad and if ℓ is assumed to be 2×10^{-8} henry, the computed resistance r is 20 to 35 ohms. This is about 500 times greater than the dc resistance of the same circuit. With this concept of the local contact circuitry $\bar{v}$ was well defined and was then possible to perform some checks of the main analytical relations presented in this paper with new measurements and with measurements previously published.

Comparison of Theory with Measurements:

(1) Voltage drop across main condenser: In Fig. 7(B) is shown the voltage drop across the main circuit condenser as a function of time for $L = 10^{-3}$ henry, $C = 1100 \times 10^{-12}$ farad and $V_0 = 220$ volts. Line (a) was measured and Line (b) was computed using Equation 4(b). The value of $\bar{v}$ used was obtained from Equation 6. Good agreement is indicated.

(2) Limiting circuit conditions for obtaining a steady arc: It was pointed out in a previous section of this paper that a steady arc can be established only if the maximum current reached during the multiple discharge period is equal to or greater than the arc initiation current. Equations 5(a) and 5(b) are expressions for the limiting circuit inductance and the limiting initial voltage respectively. Equation 5(b) was used to compute the limiting voltage for a set of circuit conditions for which measurements were made and published, Reference 3. In Table IV Column 3 of this reference measured values of the limiting voltage V_0 were presented. A comparison of measured and computed V_0 are given here as Table III. It may be pointed out that the deviations between measurements and calculations are of the order of the measured

spread in the minimum arcing current as given in the same reference, Table V.

(3) Contact activation and limiting circuit conditions for arcing: In Reference 3, it was reported that the limiting inductance for active contacts was greater by more than 2 orders of magnitude than for clean contacts. By calculation, Equation 5(a),

$$[(L)_{\text{active}}/(L)_{\text{clean}}]_{\text{Limit}} = 600.$$

It was also reported that for $C = 10^{-8}$ farad and $V_0 = 50$ volts the limiting inductance observed for active silver was between 10^{-3} and 10^{-2} henry. By calculation, Equation 5(a), the limiting inductance for the same conditions is about 4×10^{-3} henry.

(4) Contact activation and arc initiation time: According to Equation 3 the initiation time of the steady arc is directly proportional to the arc initiation current. For the same circuit conditions, therefore, active contacts should have a shorter period of arc initiation. This result seems to contradict some published observations⁴ where it was pointed out that the voltage drop into an arc was shorter for clean contacts than for activated contacts. A number of transients, furnished by the authors, were carefully examined. It was observed that in most cases

TABLE III — COMPUTED AND MEASURED LIMITING VOLTAGES FOR ESTABLISHING A STEADY ARC BETWEEN CLEAN SILVER CONTACTS*

Observed (V_0) Limit for first detectable arc	Computed (V_0) Limit
50	49
66	102
220	211
38	38
85	75
206	153
31	28
64	52
200	102
24.5	23
60	40
30	20
13	16
11.5	14
13	13
25	17

* A few observations were not included in this table. These correspond to the cases where the calculated initiation times of the arc were of the order or greater than the time of physical closure of the contacts. As expected, the observed voltages were consistently higher than the computed voltages.

⁴ L. H. Germer and J. L. Smith, Arcing at Electrical Contacts on Closure, Part, III. J. Appl. Phys., **23**, p. 553, 1952.

for *both* active and clean contacts the transient started with a rapid drop. For clean contacts this drop was directly to the steady arc voltage of the contact metal. For active contacts the drop was to a higher voltage between 15 and 32 volts followed by a gradual and irregular drop in voltage. The time of the first rapid drop was invariably between 2×10^{-9} and 4×10^{-9} second for *both* clean and active contacts over a range of circuit inductances between 0.1×10^{-6} and 48×10^{-6} henry. This time was just about the time resolution of the scope used indicating that in all cases the initiation time was less than the time resolution of the scope. This was also borne out by our calculations, Equation 3, where the longest initiation time for the conditions studied was only 1.2×10^{-9} second. The higher arcing voltage of active contacts mentioned above has been previously reported in Reference 3. It was pointed out that active contacts have arc voltages comparable with those of carbon, 19 to 30 volts.* The slow drop in voltage following the first rapid drop observed with active contacts is probably a burning off process of the activating substance on the contacts as indicated by a continuous approach of the arc voltage to that of the clean metal. It may be added that for cases where the initial circuit voltage is closer to the carbon arc voltage one should expect a smaller initial drop. This was confirmed by measurements made at 35 volts.

In connection with the experimental study of the initiation of the arc, the following concluding remark may be made. Unless the measuring apparatus has a response faster than the individual discharges at the contacts, the recorded transient will essentially be some particular average of a complex contacts transient. It should not, accordingly, be mistaken for the more fundamental and usually much faster formative transient of the arc.

The author is indebted to Dr. P. Kisliuk and Dr. L. H. Germer for much valuable discussion.

* Recent measurements by the writer on arc lamp carbon have given arc voltages as high as 43 volts.

Polyethylene Insulated Telephone Cable

By. A. S. Windeler

(Manuscript received August 25, 1953)

The physical properties of polyethylene are such as to make it attractive for many wire insulating applications, particularly in multi-conductor communications cables. This article presents certain factual information relating to new types of multi-conductor cables having extruded polyethylene insulation, and describes briefly their initial installation in working telephone plant. The literature is replete with information on the physical and chemical properties and the behavior of polyethylene, so no attempt is made to explore the quality of the material per se. Polyethylene insulation extruded in the form of both solid material and foam to impart certain desired electrical properties is discussed. In a broad sense this article may be considered as announcing an important new insulating material for telephone cables which may be expected eventually to have very extensive applications in the Bell System plant.

From almost the beginning of the art, multiconductor telephone cables have been insulated with paper, applied as a helical tape or laid down directly on the conductor in the form of pulp. Solid paper has a dielectric constant in the order of 2.5 to 3.0 but in the case of either ribbon or pulp insulation a considerable amount of air is included in the electric field surrounding the conductor so that the composite effective dielectric constant is of quite low value, usually about 1.5 to 1.6 in a typical design. In recent years, as various plastics and other polymeric materials have become available, these have been studied as competitors of paper, and polyethylene in particular now appears to have an important field of application. Polyethylene appears attractive because of its excellent electrical properties including low dielectric constant and power factor, compared with other useable plastics, and high dielectric strength. It is also highly impermeable to water or water vapor and is available in the desired quantities at a reasonable price. Additionally it is considered probable that the long term price trend will be downward.

The various electrical and mechanical characteristics^{1, 4} of polyethylene are shown in Table I, along with some of the other materials considered. Among these other materials only polytetrafluoroethylene (Teflon) and polystyrene have power factors and dielectric constants in the same low range as polyethylene. There are basic objections to both of these materials. Polytetrafluoroethylene is so expensive as to be uneconomical for this application and polystyrene in thick sections is too stiff and brittle to handle in a satisfactory manner.

The use of polyethylene in telephone cables is not altogether new but it has heretofore been confined to special types of high-frequency cable. For example, the coaxial cable and video pair have polyethylene disc insulation and strip-and-string insulation respectively (see Fig. 1). These cables were designed for low attenuation in the megacycle range and the use of polyethylene or a similar low power factor material was a necessity. A low power factor is of lesser importance in the carrier systems for which the multipair cables are used and the polyethylene insulation on these cables must show other advantages in order to prove in.

Polyethylene insulation is applied to the wire by an extrusion process; the insulation may be either solid or expanded depending on the application. Generally the polyethylene is supplied as granules previously compounded with an antioxidant. In the case of solid polyethylene insulation the granules, in which the pigment has been incorporated, are fed into the extruder and formed on the conductor as a uniform close fitting tube of insulation.

TABLE I — CHARACTERISTICS OF INSULATING MATERIALS

	Polyethylene	Plasticized Polyvinyl- Chloride	Polystyrene	Polytetra- fluoroethylene (Teflon)	Polyamide (Nylon)
Density—gms/cc.	0.92	1.2-1.4	1.06	2.2	1.09
Tensile strength— psi	1400-2000	1500-3000	500-9000	1500-2500	7000
Elongation %	600	200-450	2-5	100-200	100-200
Water absorption % in 24 hrs.	<0.01	0.4-0.65	<0.05	Nil	0.4-2.0
Diel. strength, RMS volts/mil $\frac{1}{8}$ " thickness*	400-500	300-700	500-700	400-500	400
Power factor, 1- 300 kc	0.0002	0.09-0.16	0.0002	0.0002	0.04-0.2
Diel. constant, 1- 300 kc	2.3	3.5-5.0	2.5-2.6	2.0	3.5-8

* Dielectric strengths are greater for thinner sections—for example, in 14 mil thicknesses polyethylene has a dielectric strength of approximately 2500 RMS volts per mil.

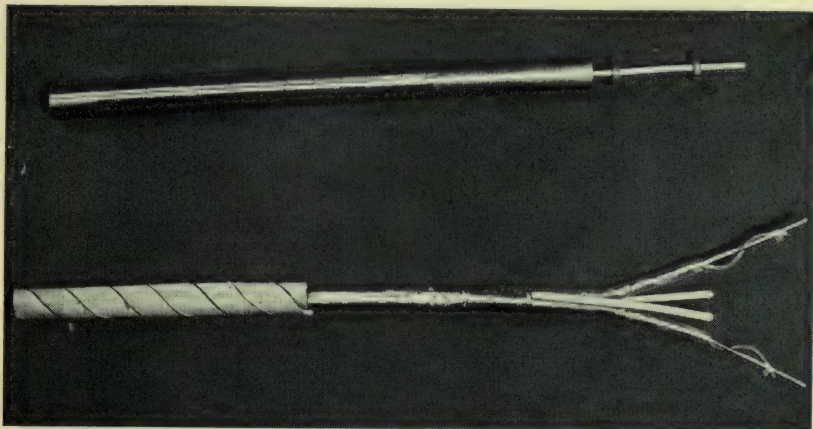


Fig. 1—(a) Polyethylene disc insulated coaxial. (b) Polyethylene ribbon and string insulated video pair. Note expanded polyethylene interstice fillers.

The idea of using spongy or foamed hydrocarbons as conductor insulation is not new. A British patent issued in 1930 contemplates such a structure and numerous United States patents of more recent dates cover various aspects of cellular hydrocarbon insulation. The problem is one of forming the cylinder of aerated plastic in an extrusion process operating at high speed and producing a closely controlled uniform covering having precise physical and electrical properties. The original development work was carried on by F. B. Lyons of the Western Electric Company in cooperation with the author. This early work demonstrated that material having the desired range of properties could be applied in a continuous extrusion process and subsequent work has shown that the necessary control of properties and speed of extrusion can be achieved.

The expansion of polyethylene is accomplished by methods similar to those employed in the making of many of the numerous polymer and rubber "foams". The process used to produce the cellular polyethylene involves the addition at the extruder of a chemical blowing agent which decomposes under heat and releases nitrogen gas. By proper mixing and process control this nitrogen gas can be entrapped in the polyethylene in the form of very small discrete bubbles, thus achieving the cellular structure shown in Fig. 2. It is interesting to note that the foamed plastic tends to form a desirable "skin" of solid material on the inner surface over the conductor, Fig. 2(a).

Various degrees of expansion can be achieved as required by varying the amount of blowing agent and by other means. The degree of expansion, or percent entrapped gas, can be determined readily by weigh-

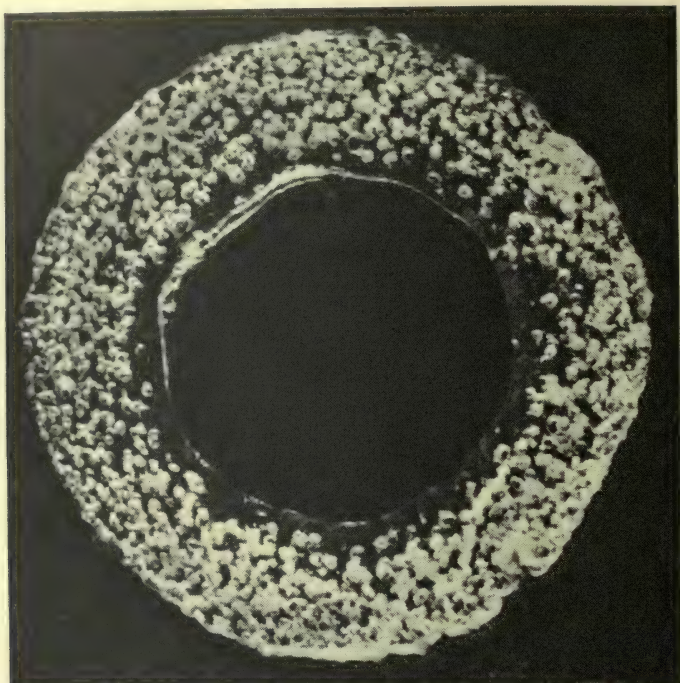


Fig. 2(a)—Cross section of expanded polyethylene insulation from 19 gauge conductor—35 per cent air. Magnified 75 times.

ing a sample of insulation, with the conductor removed, on an analytical balance. The inside and outside diameter of the cylinder of insulation and the density of solid polyethylene are required to complete the determination.

The composite dielectric constant obviously varies with the degree of expansion. In a coaxial configuration this effect is calculable from the formula for the dielectric constant of a mixture,⁶ the relation being given by the following:

$$\frac{\epsilon - \epsilon_p}{3\epsilon} = V \frac{(\epsilon_a - \epsilon_p)}{(\epsilon_a + 2\epsilon)}$$

where ϵ = composite dielectric constant

ϵ_p = dielectric constant of polyethylene = 2.26

ϵ_a = dielectric constant of added material, in this case 1 for air

V = volume fraction = $\frac{\text{percent air}}{100}$

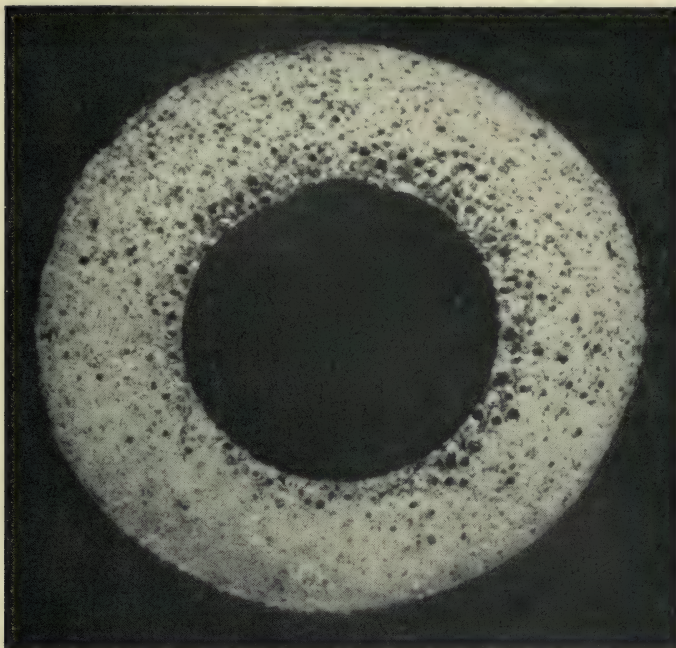


Fig. 2(b)—Cross section of expanded polyethylene insulation from 19 gauge conductor—55 per cent air. Magnified 75 times.

The upper curve in Fig. 3 was determined in this manner. However, in multipair cable the dielectric constant is not amenable to calculation and must be determined experimentally.² The empirical relation between the percent air in the insulation and the dielectric constant in the cable for a typical design is shown in the lower curve in Fig. 3. Effective dielectric constants as low as 1.40 have been achieved in expanded polyethylene cables.

Solid polyethylene insulated cables are more costly for given transmission characteristics than those insulated with paper or pulp and, therefore, are restricted to special uses where a system saving can be obtained in spite of the higher first cost. One such use is for small aerial toll cables in rural areas where, with paper-insulated cable, maintenance costs are likely to be high. There are several factors which tend to increase the maintenance costs in such cables. For instance, in small isolated cables lightning troubles are common because of the high sheath resistance.³ While the incidence of sheath breaks from other causes is usually neither more nor less than in other aerial cables, maintenance is more difficult and costly because of inaccessibility. Maintaining gas pressure on these cables is expensive for the same reason.

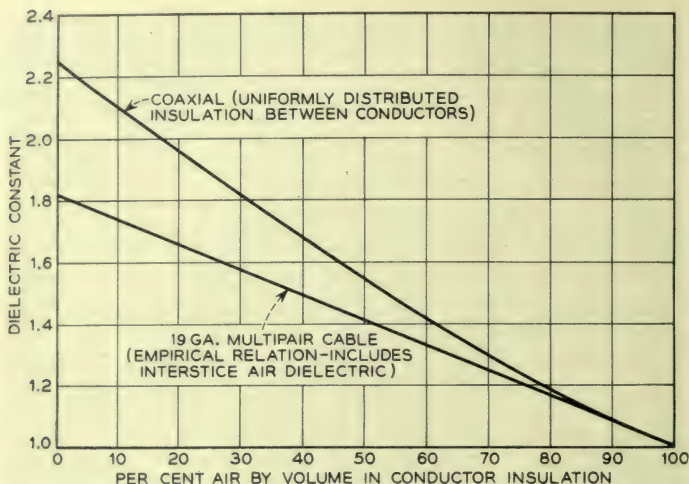


Fig. 3—Change in dielectric constant of coaxial cable and multipair cable with degree of expansion of polyethylene conductor insulation.

In the case of a sheath break with no insulation damage in polyethylene-insulated cables, there is no interruption in service due to entrance of moisture into the cable and repairs can be made on a routine maintenance basis. Gas pressure maintenance is unnecessary and its omission effects appreciable annual savings. Because these cables usually contain not more than 51 pairs the first cost penalty for the use of solid polyethylene, in terms of cents per foot, is small. Small aerial toll cables in rural areas, therefore, are the most promising candidates for solid polyethylene insulation.

The solid polyethylene insulation development has progressed to the point where several cables have been made for field trials. The first of these was at Cooperstown, New York, where approximately four miles of 26-pair 19-gauge cable was installed aerially. The sheath on this cable was a composite of aluminum and polyethylene commonly known as "alpeth" sheath. The Cooperstown cable (see Fig. 4) was one of the first to be installed by the pre-lashing method which has been developed as a means for effecting economies in placing aerial cable.

A second trial installation of solid polyethylene insulated cable was made between Trout Lake and St. Ignace, Michigan, a distance of twenty-eight miles. This cable contained 51 pairs of 19 gauge and was also covered with alpeth sheath. Since that time solid polyethylene cables have been installed in other locations where the anticipated maintenance savings were believed to justify the higher first cost.

Development of expanded polyethylene insulation has been carried

on concurrently with that of solid polyethylene. In expanded polyethylene insulation the cost, as would be expected, varies with the degree of expansion. There are two reasons for this: first, as the proportion of gas is increased, less polyethylene is used in the insulation, and second, because of the lower dielectric constant, the cable can be made smaller for the same attenuation, resulting in savings in sheathing materials. An initial trial installation of about nineteen miles of 51-pair 19-gauge expanded polyethylene insulated cable has been completed between Grandville and Zeeland, Michigan. This route is to be developed for N carrier, and since recent Systems' studies have indicated the lower overall costs will result if low attenuation cable is used, the Grandville-Zeeland cable was designed for a capacitance of 0.066 microfarad per mile rather than the 0.084 microfarad per mile capacitance for which earlier polyethylene-insulated cables were designed. The Cooperstown, Trout Lake and Grandville cables are illustrated in Fig. 5.

It is of interest to compare polyethylene-insulated cables with paper-insulated cables having the same voice frequency attenuation. Some of the more important characteristics are shown in Table II. It will be

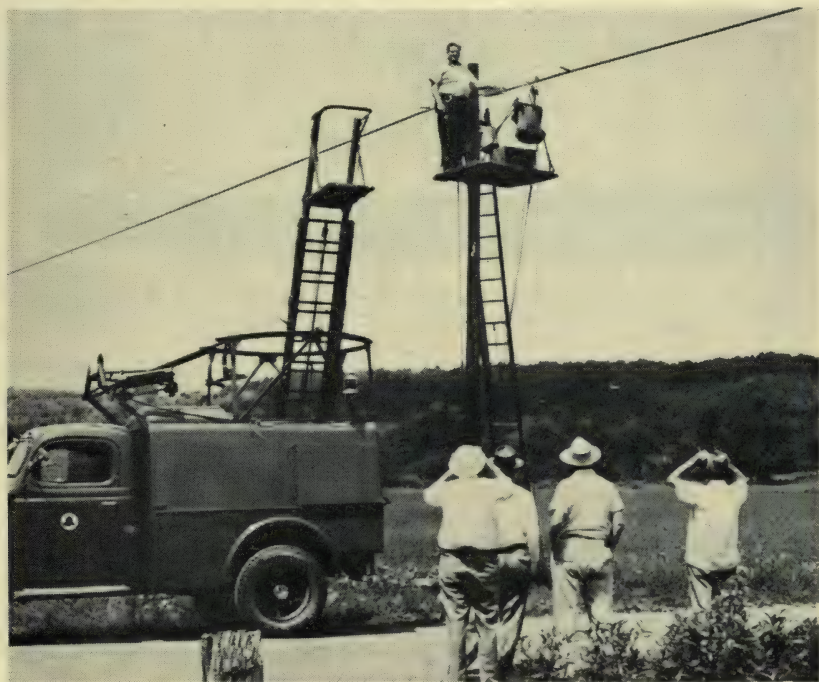
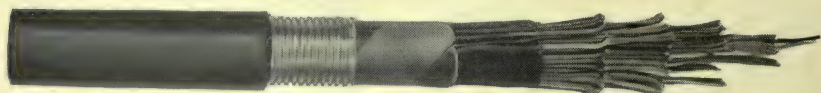


Fig. 4—Installation of Cooperstown-Cherry Valley (N. Y.) cable.

noted that the polyethylene-insulated cables excel in dielectric strength. The values shown in Table II are the test voltages that all the reels of cable were required to withstand between conductors and from core to sheath. These voltages are naturally made lower than the inherent dielectric strength of the insulation to allow for minor manufacturing irregularities and thus avoid excessive rejections. Some idea of the inherent dielectric strength of solid and expanded polyethylene insulation can be obtained from tests on short samples of insulated conductor immersed in water. In Fig. 6 the inherent dielectric strength obtained in this manner for conductors with a 14-mil wall is plotted against percent air in the insulation. It will be noted that the dielectric strength falls off rapidly as the polyethylene is expanded.

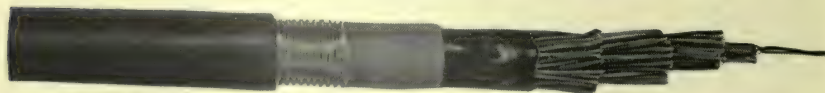
The capacitance unbalance to ground, or difference in direct capacitances of the wires of a pair to ground, is a rough indication of one important factor in the susceptibility of cable circuits to noise and interference. The electrical disturbances which cause noise in the cable may come from atmospheric static, radio stations, power lines, or from telephone plant sources. The former energize the cable circuits via the surface transfer impedance⁵ of the sheath, or by way of an open wire tap, the latter via all of the conductors in the cable. For this reason the two wires of a pair should have nearly equal capacitance to the surrounding pairs and to the sheath. To achieve this condition the cylinders of insulation on the two wires of a pair must be alike in size and dielectric



GRANDVILLE-ZEELAND (MICH.) PROJECT
51 PAIRS NO. 19 GAUGE - EXPANDED POLYETHYLENE INSULATION



COOPERSTOWN-CHERRY VALLEY (N.Y.) PROJECT
26 PAIRS NO. 19 GAUGE - SOLID POLYETHYLENE INSULATION



TROUT LAKE-ST. IGNACE (MICH.) PROJECT
51 PAIRS NO. 19 GAUGE - SOLID POLYETHYLENE INSULATION

Fig. 5—Polyethylene insulated multipair cables.

TABLE II — COMPARATIVE DATA ON 51-PAIR 19-GAUGE CABLES

Type of Insulation	Inside Diameter of Sheath	Mutual Cap. μf per Mile	Attenuation* DB per Mile—70°F			Average Unbalanced to Ground $\mu\text{f}/1500'$	Minimum Dielectric Strength—KV	
			1 kc	60 kc	150 kc		Cond. to Cond.	Core to Sheath
Solid polyethylene (Trout Lake project)	0.92"	0.082	1.24	4.51	6.86	104	10	10
Paper (std. CNB).	0.80"	0.084	1.26	4.94	7.79	310	0.7	1.4
Expanded polyethylene (Grandville project)	0.97"	0.066	1.11	3.92	5.96	118	3	10
Paper (std. DNB).	0.94"	0.066	1.11	4.05	6.40	240	0.7	1.4

* Computed from primary constants.

constant. The low value of capacitance unbalance obtained with solid polyethylene insulation is a result of the remarkable uniformity with which this insulation is extruded. The value of 104 micro-microfarads per 1500 feet for the capacitance unbalance to ground is on the average only 0.4 per cent of the direct capacitance of either wire to ground.

The other important factor in unbalance is the uniformity of the twisting, that is, the extent to which wire and mate form symmetrical helixes around the center line of the pair. Polyethylene-insulated pairs appear to be better in this respect for reasons which are not very obvious. The fact that the polyethylene forms firm tubes of insulation of equal size on both conductors of the pair probably is a factor in achieving this uniform twisting.

The precision or accuracy with which the length of pair twist is maintained is also better with the solid polyethylene insulated conductors. If an adequate number of different twist lengths is used in the cable precision twisting is an advantage since the cross talk between adjacent pairs is reduced over that which exists in the less precisely twisted paper insulated cables. The cross talk between pairs with like lengths of twist is kept to a low value by designing the cable so that such pairs are widely separated.

The carrier-frequency attenuation of polyethylene-insulated cable is substantially lower than that of paper-insulated cable which has approximately the same voice-frequency attenuation. In the case of solid polyethylene cable there are two reasons for this lower carrier-frequency attenuation — the inductance is higher and the conductance is lower than those in comparable paper-insulated cable. The higher inductance

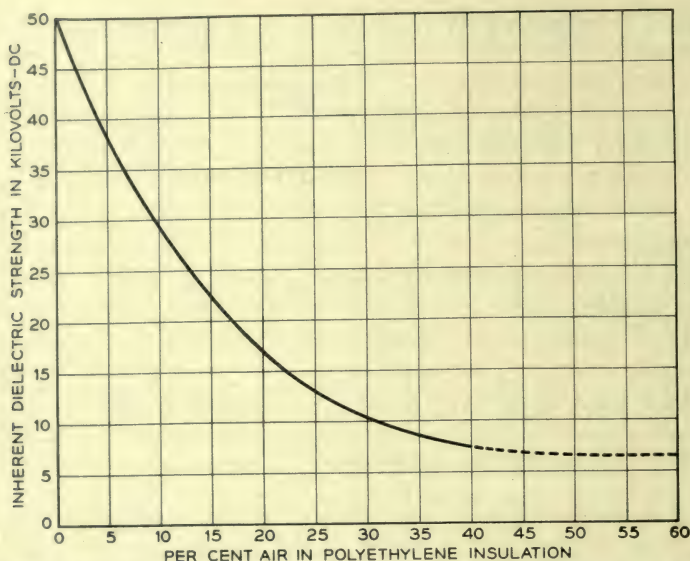


Fig. 6—Inherent dielectric strength of expanded polyethylene versus degree of expansion. 19-gauge 0.064-inch D.O.D. conductors. Short samples immersed in water.

is a result of the greater separation between the wires of a pair and the lower conductance is a result of the lower power factor of polyethylene compared to that of paper. The voice frequency attenuation is relatively independent of inductance and conductance.

While the electrical characteristics of polyethylene cables are superior to those of paper cables, the higher first cost of cables insulated with solid polyethylene has been a deterrent to their widespread use. This higher first cost is inherent in solid polyethylene because, in addition to the higher cost of polyethylene as compared to paper, the cables must be larger for the same voice frequency attenuation. It will be noted that the 51-pair Trout Lake-St. Ignace cable is 15 per cent larger in diameter than the comparable paper cable. The larger size is necessary because of the higher effective dielectric constant which is approximately 1.80 in solid polyethylene cable as compared to 1.60 in a typical paper cable. The effective dielectric constant is higher in the case of solid polyethylene insulation because of the lesser amount of air space which can be incorporated in the dielectric between wires.

As was illustrated in Fig. 3, the dielectric constant can be decreased by expanding the polyethylene. A value as low as 1.40 has been attained experimentally. The effective dielectric constant of the Grand-

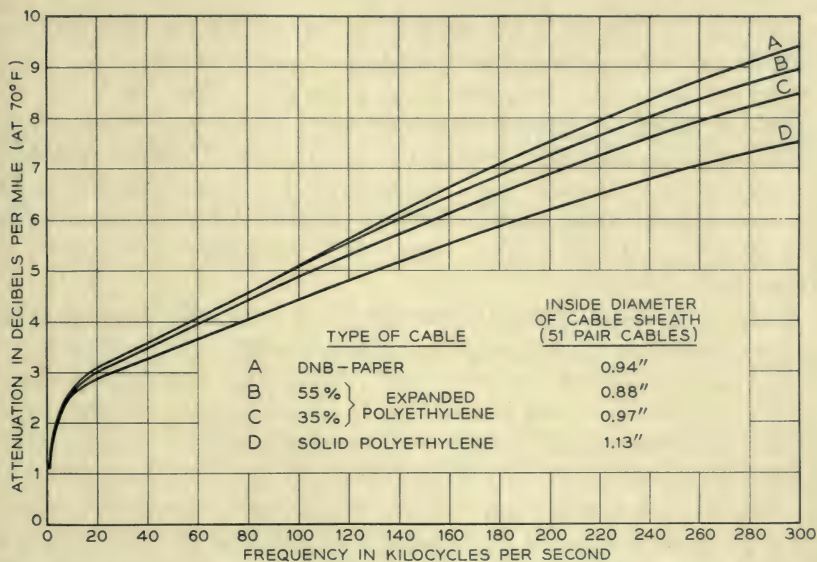


Fig. 7—Attenuation versus frequency. 19-gauge cables having $0.066 \mu\text{f}$ per mile capacitance.

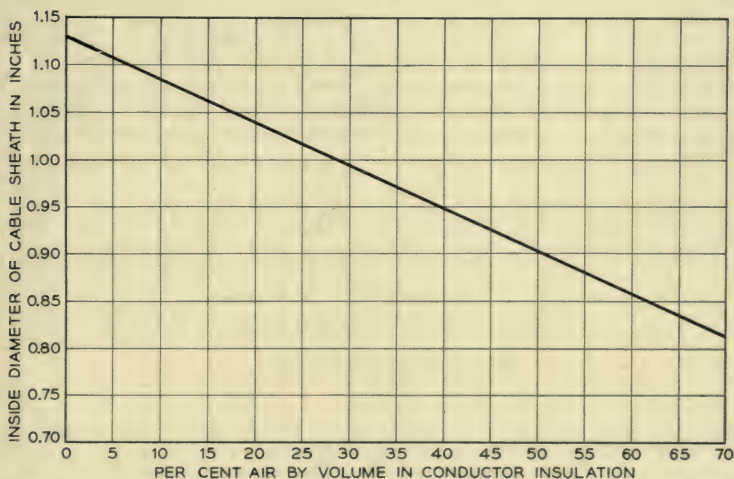


Fig. 8—Change in cable diameter with degree of expansion of polyethylene conductor insulation. 19-gauge 51-pair cable, $0.066 \mu\text{f}$ per mile capacitance.

ville-Zeeland cable was 1.54, obtained by expanding the conductor insulation to the point that 35 per cent of the volume was gas.

The curves on Fig. 7 represent the attenuation over the N carrier-frequency range for a paper-insulated cable and three polyethylene-insulated cables all designed to have a capacitance of 0.066 microfarads per mile and a voice frequency attenuation of 1.1 db per mile. The lower carrier-frequency attenuation of the polyethylene cables is evident. The size of the polyethylene cables varies with the degree of expansion as illustrated in Fig. 8.

Since the dielectric strength decreases as the degree of expansion increases, the saving in first cost must be balanced against the value of the reduction in reliability. As mentioned before, the Grandville-Zeeland project was the first trial installation of expanded polyethylene cable. A very moderate degree of expansion was chosen for this project. However, as more experience with expanded polyethylene is gained, it should be possible, for a given application, to determine the degree of expansion which strikes the optimum balance between dielectric strength and dielectric constant, giving proper weight to the mechanical properties and to cost factors.

REFERENCES

1. Modern Plastics Encyclopedia.
2. Maupin, J. T., Measurements in Multipair Cables. Bell Sys. Tech. Jl., **30**, pp. 652-667, July, 1951.
3. Horn, F. W., and H. B. Ramsey, Bell System Sheath Problems and Designs. A.I.E.E., Trans., **70**, Part I, pp. 1811-1816, 1951.
4. Shockleton, W. J., General Characteristics of Polyethylene. A.I.E.E., Trans., **64**, pp. 912-916, Dec., 1945.
5. Schelkunoff, S. A., Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields. Bell Sys. Tech. Jl., **13**, p. 532, Oct., 1934.
6. Boettcher, C. J. F., Recueil des Travaux Chimiques des Pays Bas, **64**, p. 47, 1945.

Abstracts of Bell System Technical Papers*

Not Published in this Journal

ANDERSON, P. W.,¹ AND P. R. WEISS⁶

Exchange Narrowing in Paramagnetic Resonance, *Revs. Mod. Phys.*, **25**, pp. 269-276, Jan., 1953.

In this paper the problem of the line shape in paramagnetic resonance when large exchange interaction is present is discussed from the standpoint of a simplified mathematical model. The mathematical model can be called the model of "random frequency modulation": It is assumed that the atom absorbs a single frequency, which varies over a distribution determined by the dipolar local fields, but that this frequency varies randomly in time at a rate determined by the exchange interactions. The predicted line shape in the case in which exchange is large is of resonance type in the observable center of the line, but falls off more rapidly in the wings. This line shape has been verified experimentally in a number of cases. This conclusion seems quite independent of any assumption about the type of random frequency modulation, etc.

The quantitative conclusions are reached in the following way: It is suspected, since the exchange motion is the superposition of the effects of a number of neighbors which is not particularly small, that a good approximation to the modulation function is Gaussian noise with a Gaussian spectrum. This, of course, is what would result from the superposition of a large number of rather small effects. Under this assumption both the second moment (which is independent of exchange) and the fourth moment of the line shape can be calculated. This kind of modulation is the simplest one which does give a finite fourth moment; a Markoffian, or "jump", type of modulation, which might seem more reasonable at first, does not. These moments are then compared with the moments computed by Van Vleck [*Phys. Rev.* **74**, 1168 (1948)] to fix the two adjustable parameters, mean square frequency, and average rate of change of frequency, of the theory.

The result as to line breadth, which is essentially

$$\Delta \cong \frac{\langle (\Delta\omega^2) \rangle_{AV} \text{ dipole-dipole}}{J/h}$$

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

¹ Bell Telephone Laboratories.

⁶ Rutgers University, New Brunswick, N. J.

if J is the exchange integral, can be compared with observed line breadths by estimating J from Curie-Weiss constants for a number of materials. The results are quite satisfactory if the theory is extended in two ways: (a) When the exchange frequency is larger than the resonance frequency, it can be shown that the off-diagonal elements of the dipolar interaction must be included, leading to a line-width larger by a factor of roughly $10/3$; (b) in a number of cases hyperfine and Stark splitting is contributing importantly to the width.

The good agreement with experiment in the cases we have investigated leads us to believe that a quantitative approach to the paramagnetic resonance line breadth problem, using only the already known concepts of dipolar interaction, exchange narrowing, and fine structure splitting, will probably explain all the observed phenomena.

ANDRUS, J., see J. K. GALT.

ARNOLD, S. M., see MISS S. E. KOONCE.

BENEDICT, T. S.,¹ AND W. SHOCKLEY¹

Microwave Observation of the Collision Frequency of Electrons in Germanium, Letter to the Editor. *Phys. Rev.*, **89**, pp. 1152-1153, Mar. 1, 1953.

BIES, F. R.¹

Attenuation Equalizers, *Audio Eng. Soc. J.*, **1**, pp. 125-136, Jan., 1953.

In all systems there are components which attenuate some frequencies to a greater extent than others, and attenuation equalizers are usually required to correct the overall gain-frequency characteristic. This paper will deal with the types of attenuation equalizer that are found most useful, the performance that they display, and a chart method of computing their insertion loss.

BOZORTH, R. M.¹

Permalloy Problem, *Revs. Mod. Phys.*, **25**, pp. 42-48, Jan., 1953 (Monograph 2102).

In attempting to explain the unusual magnetic properties of the iron-nickel alloys, single crystals of alloys containing 35 to 100 per cent nickel were prepared, and measurements made of the magnetic crystal anisotropy and magnetostriction as dependent on cooling rate. It is confirmed that there is a large effect of cooling rate on the anisotropy in the region near FeNi_3 , but the experiments show also a substantial effect between 50 and 85 per cent nickel. Two magnetostriction constants, λ_{100} and λ_{111} , were measured on the

¹ Bell Telephone Laboratories.

same crystals. The effect of cooling rate on magnetostriction was found to be substantial only in the composition range 70 to 80 per cent nickel. When the specimens are quenched, λ_{111} goes through zero for a nickel content just below 80 per cent nickel, a composition very close to that for highest permeability. This is understandable because the magnetostrictive strain caused by movement of the boundary between two domains, each magnetized spontaneously in a [111] direction, depends on λ_{111} alone. The same physical picture predicts that near 45 percent nickel, where [100] is the direction of easiest magnetization and λ_{100} goes through zero, the permeability versus composition curve should again have a maximum. Such a maximum is known to exist, and initial permeabilities as high as 15,000 have been observed. Although simple theory suggests that domain-rotation should occur in very weak fields when the crystal anisotropy is very small (75 per cent nickel in quenched alloys), nevertheless, rotation involves magnetostrictive strains which prevents μ_0 from becoming infinite. Internal poles are also likely to be formed. In slowly cooled alloys the anisotropy is zero at about 63 per cent nickel; here there are random strains caused by magnetostriction and possibly also by atomic ordering. The principal changes in magnetic properties with composition are explained in terms of the crystal anisotropy and magnetostriction, and their change with heat treatment.

BOZORTH, R. M.¹

Behavior of Magnetic Materials, Am. J. Phys., **21**, pp. 260-266, Apr., 1953 (Monograph 2105).

This is a review of recent work in which the atomic theory of ferromagnetism and the domain theory of magnetization are applied to new materials.

BOZORTH, R. M., see H. J. WILLIAMS.

BURNS, R. M.¹

Science and Scientists in Telecommunications, Electrochem. Soc. J., **100**, pp. 90C-94C, Apr., 1953.

COLLEY, R. H., see G. Q. LUMSDEN

COY, J. A.,¹ AND E. K. VAN TASSEL¹

Type-O Carrier Telephone, Elec. Eng., **72**, pp. 418-423, May, 1953.

The Type-O carrier is an economical short-haul carrier system especially suitable for use under 150 miles. It fulfills the same purpose for open-wire lines as the Type-N carrier system does in cable routes. Numerous laboratory tests have indicated that good service standards have been maintained in spite of its low cost.

¹ Bell Telephone Laboratories.

FINE, M. E.¹

Magnetomechanical Effects in an Antiferromagnet, CoO, Revs. Mod. Phys., **25**, p. 158, Jan., 1953.

FINE, M. E.¹

Elasticity and Thermal Expansion of Germanium between -195 and 275°C, J. Appl. Phys., **24**, pp. 338-340, Mar., 1953 (Monograph 2069).

Young's moduli (E) of the directions (100) and (111) and the shear modulus (G) for (100) were determined in germanium from -195 to 255, 275, and 140°C, respectively. From these moduli, the elastic parameters, the compressibility, and Poisson's ratio were calculated. The thermal expansion was measured from -196 to 275°C.

GALT, J. K.,¹ ANDRUS¹, J. AND H. G. HOPPER¹

Motion of Domain Walls in Ferrite Crystals, Revs. Mod. Phys., **25**, pp. 93-97, Jan., 1953.

GRAY, A. N.²

Development of Electroformed Copper-steel Wire, Wire and Wire Products, **28**, pp. 166-168, 218-219, Feb., 1953.

The author traces the development and the functioning of the apparatus in the successful efforts to produce a satisfactory copper coated steel wire used for telephone drop wire.

GROTH, W. B.¹

Principles of Tape-to-Card Conversion in the AMA System, A.I.E.E. Trans. Commun. and Electronics Sect., **5**, pp. 42-52, Mar., 1953.

HAYNES, J. R.,¹ AND J. A. HORNBECK¹

Temporary Traps in Silicon and Germanium. Letter to the Editor. Phys. Rev., **90**, pp. 152-153, Apr. 1, 1953.

HEIDENREICH, R. D., see E. A. NESBITT.

HOGAN, C. L.¹

Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications, Revs. Mod. Phys., **25**, pp. 253-262, Jan., 1953.

¹ Bell Telephone Laboratories.

² Western Electric Company, Inc.

HOPPER, H. G., see J. K. GALT.

HORNBECK, J. A., see J. R. HAYNES.

HUGHES, W. T.¹ AND J. J. LANDER¹

Vacuum Tube Electrometer Amplifier, *Rev. Sci. Instr.* **24**, pp. 331-332, Apr., 1953.

KERSTA, L. G.¹

Interesting Property of Certain Conductive Rubbers. Letter to the Editor. *J. Polymer Sci.*, **10**, pp. 447-448, Apr., 1953.

KOCK, W. E.¹

Acoustic Gyrator, *J. Am., Acoust. Soc.* **25**, p. 575, May, 1953.

KOCK, W. E.¹

Acoustic Gyrator, *Arch. Elektr. Übertragung*, **7**, p. 106, Feb., 1953.

A gyrator for acoustic waves is described which is the analog of the ferrite gyrator for microwaves described by C. L. Hogan. The required non-reciprocal rotation of the plane of polarization of transverse acoustic waves propagating in a tube is accomplished by rotating the tube at high speed.

KOCK, W. E.¹

Ähnlichkeit zwischen Vokalformanten und Formanten von Musikinstrumenten (Similarity Between Vowel Formants and the Formants of Musical Instruments), (in German), *E.T.Z.*, **74**, p. 166, Mar. 1, 1953.

KOHMAN, G. T., see M. C. WOOLEY.

KOONCE, MISS S. E.,¹ AND S. M. ARNOLD¹

Growth of Metal Whiskers. Letter to the Editor, *J. Appl. Phys.*, **24**, pp. 365-366, Mar., 1953.

LANDER, J. J., see W. T. HUGHES.

LUKE, C. L.¹

Photometric Determination of Antimony in Lead Using the Rhodamine B Method, *Anal. Chem.*, **25**, p. 674, Apr., 1953.

¹ Bell Telephone Laboratories.

LUMSDEN, G. Q.,¹ AND R. H. COLLEY¹

Review of American Standard Fiber Stresses of Wood Poles, Standardization, 24, pp. 114-117, Apr., 1953.

McKAY, K. G.¹

Crystal Conduction Counter, Physics Today, 6, pp. 10-13, May, 1953.

Development of the crystal counter (essentially an ionization chamber that is solid instead of being filled with gas) has progressed in the last few years, but is still delayed for lack of better understanding of crystals and their electrical behavior.

McMAHON, W., see M. C. WOOLEY.

MASON, W. P.¹

Rotational Relaxation in Nickel at High Frequencies, Revs. Mod. Phys., 25, pp. 136-139, Jan., 1953.

Measurements of the ΔE effect and the decrement made by Bozorth, Mason, and McSkimin and by Johnson and Rogers are compared with that expected from a calculation of domain wall relaxations for a distribution of domain sizes as determined by the optical measurements of Williams and Walker. At low frequencies the agreement is good but at high frequencies a second relaxation region is indicated. It is shown that this region is consistent with a domain rotation relaxation and introducing this effect, a good agreement is obtained between theory and experiment for the entire frequency range.

MATHY, P. T.²

Clever Conveyorization Avoids Handling Headaches, Am. Mach., 97, pp. 140-143, May 11, 1953.

MORRISON, J.¹

Controlled Gas Leak, Rev. Sci. Instr., 24, pp. 230-231, Mar., 1953.

NESBITT, E. A.,¹ AND R. D. HEIDENREICH¹

Physical and Magnetic Structure of the Mishima Alloys, Rev. Mod. Phys., 25, pp. 322-323, Jan., 1953.

¹ Bell Telephone Laboratories.

² Western Electric Company, Inc.

PEARSON, G. L.,¹ AND M. TANENBAUM¹

Magnetoresistance Effect in InSb. Letter to the Editor, Phys. Rev., **90**, p. 153, Apr. 1, 1953.

PUGH, S. G.⁵

Southern Bell Switches to Chemical Brush Control, Elec. World, **139**, pp. 122-123, Mar. 23, 1953.

If properly planned and executed, chemical control costs less in time and money and does a better job.

RALSTON, R. W.⁴ AND B. D. WICKLINE⁴

Television Coverage of the National Political Conventions, A.I.E.E. Trans. Commun. and Electronics Sect. **5**, pp. 1-14, Mar., 1953, and Elec. Eng., **72**, pp. 383-389, May, 1953.

The first large-scale television coverage of both national political conventions occurred last year in Chicago and this presented many new problems to the telephone company of that city. Special video conductors and amplifiers were used in eight of the 19 channels to the amphitheatre and microwave facilities for the rest.

SHIVE, J. N.¹

Properties of Germanium Phototransistors, J. Opt. Soc. Am., **43**, pp. 239-244, Apr., 1953 (Monograph 2103).

This paper describes, summarizes, and compares the properties of three photoelectric devices, namely: point contact phototransistors, p - n junction phototransistors, and n - p - n junction multiplier phototransistors, which have resulted from the prosecution of the transistor program at Bell Telephone Laboratories. The first of these devices is characterized by a comparatively high dark current and a quantum yield of 3 or 4 electrons per quantum. The second has a dark current in the microampere range and a quantum yield of approximately unity. The n - p - n device has a sensitivity corresponding at best to a quantum yield of several hundred electrons per quantum. All these devices have long-wave thresholds around 1.8 microns. The structures lend themselves readily to miniature encapsulation.

SHOCKLEY, W., see T. S. BENEDICT.

¹ Bell Telephone Laboratories.

⁴ Illinois Bell Telephone Company.

⁵ Southern Bell Telephone Company.

SLADE, F. D.²

Mechanized Billing of AMA Toll Messages, A.I.E.E., Trans. Commun. and Electronics, **6**, pp. 175-182, May, 1953.

SLEPIAN, D.¹

On the Number of Symmetry Types of Boolean Functions on n Variables, Can. J. Math., **5**, pp. 185-193, 1953.

TANENBAUM, M., see G. L. PEARSON.

VAN SICLEN, H. E.³

What We Did to Cut Costs of Finishing Telephone Woodwork, Ind. Finishing, **29**, pp. 44-52, Apr., 1953.

VAN TASSEL, E. K., see J. A. COY.

WILLIAMS, H. J.,¹ AND R. M. BOZORTH¹

Magnetic Study of Low Temperature Transformation in Magnetite, Rev. Mod. Phys., **25**, pp. 79-80, Jan., 1953.

WOOD, E. A.¹

Simple Attachment for Low Temperature Use of an X-Ray Diffraction Camera, Rev. Sci. Instr., **24**, pp. 325-326, Apr., 1953.

WOOLEY, M. C.,¹ G. T. KOHMAN,¹ AND W. McMAHON,¹

Polyethylene Terephthalate — Its Use as a Capacitor Dielectric, A.I.E.E. Trans. Commun. and Electronics Sect., **5**, pp. 33-37, Mar., 1953 (Monograph 2125).

The steady increase in the severity of operating conditions for capacitors and the need for more diverse characteristics has spurred the search for new and better capacitor dielectrics. The synthetic plastics industry, which is the source of a number of useful dielectric materials, has recently produced a new and promising dielectric in film form known as polyethylene terephthalate or "Mylar." This material is unusually strong, has a high softening point, and is available in very thin films which makes it especially suitable for capacitor insulation. The electrical characteristics of capacitors wound

¹ Bell Telephone Laboratories.

² American Telephone and Telegraph Company.

³ Western Electric Company, Inc.

with Mylar film are likewise promising. As compared to mineral-oil-impregnated paper capacitors, unimpregnated Mylar capacitors exhibit higher dielectric strength, higher insulation resistance, and can be operated at higher ambient temperatures. Their loss characteristics are comparable with those of impregnated paper and the capacitance stability over the usual range of ambient temperatures approaches that of mica capacitors. Mylar is relatively nonhygroscopic and tests indicate that for moderate atmospheric conditions capacitors made from it do not require additional moisture protection. The film can be metallized readily by current techniques and, when used in metallized capacitors, appears to possess advantages over metallized paper in several respects.

WRIGHT, S. B.¹

Higher Frequencies for Ground-Air Communications, Air Univ. Quart. Rev., 5, No. 4, pp. 60-70, 1952-1953.

¹ Bell Telephone Laboratories.

Addendum to Delay Curves for Calls Served at Random

By JOHN RIORDAN

B. S. T. J.

January 1953

Pages 100 to 119

I owe the following remarks, which help to complete the record, to Emile Vaultot:

1. The Erlang formula for delay with order of arrival service, for the proof of which reference has been made to a paper by E. C. Molina, was proved earlier by E. Vaultot (*Application du Calcul des Probabilités a l'exploitation telephonique. Revue Générale de l'Electricité*, 16, pp. 411-418, 1924). Indeed his seems to be the first proof.

Also, the associated Erlang C function $C(c, a)$, for which I said there was no extensive tabulation, is tabulated for $n = 1$ (1) 139 and an extensive but irregular set of a 's by Arne Jensen (*Moe's Principle*, Table III, Copenhagen Telephone Co. Copenhagen, 1950). Also the recurrence relation for this function given in a footnote has previously been given by Conny Palm (*Väntetider Vid Slumpvis Avverkad Kö*, Tekniska Meddelanden Fran. Kungl. Telegrafstyrelsen, Specialnummer för Teletrafikteknik, pp. 109. Stockholm, 1946, see p. 43).

2. The extensive treatment of delay by Conny Palm, just mentioned, includes a section on random service (section 4); it may be noticed that this is dated May 15, 1946, which is only a few months after Vaultot's article on the same subject (Jan. 28, 1946), and of course is an independent development.

I owe the following to my colleague S. O. Rice. Pollaczek, in the *Comptes Rendus* paper mentioned, has given an integral effectively for what I have called $F(u)$. Rice has put this in a slightly different form adapted to numerical computation and has obtained the following results for $F(u)$

$v = u(1-\alpha)$								
α	1	2	4	6	8	10	12	14
0.8					0.0079	0.0039	0.0020	0.0011
0.9	0.2866	0.1388	0.0471	0.0198	0.0094	0.0049	0.0026	0.0015

Comparison with the tables of the papers shows a satisfying agreement and substantiates the conjecture that approximation by a relatively small number of exponentials is sufficient.

JOHN RIORDAN

Contributors to this Issue

REUBEN E. ALLEY, JR., B.A., University of Richmond, 1938; E.E., Princeton University, 1940; Ph.D., Princeton University, 1949. Massachusetts Institute of Technology, Radiation Laboratory, 1942 and 1943; University of Richmond, 1948-51; Bell Telephone Laboratories, 1952-53. While at Bell Laboratories, Mr. Alley was engaged in investigations of the magnetic properties of ferrites, with particular interest in frequencies above 20 megacycles. He recently accepted an appointment at the University of Richmond as an associate professor of physics. Member of the American Physical Society, A.I.E.E., I.R.E., Phi Beta Kappa and Sigma Xi.

M. M. ATALLA, B.S., Cairo University, 1945; M.S., Purdue University, 1947; Purdue University, Ph.D., 1949; Studies at Purdue undertaken as the result of a scholarship from Cairo University for four years of graduate work. Bell Telephone Laboratories, 1950-. For the past three years he has been a member of the Switching Apparatus Development Department, in which he is supervising a group doing fundamental research work on contact physics and engineering. Current projects include fundamental studies of gas discharge phenomena between contacts, their mechanisms, and their physical effects on contact behavior; also fundamental studies of contact opens and resistance. In 1950, an article by him was awarded first prize in the junior member category of the A.S.M.E. He is a member of Sigma Xi, Sigma Pi Sigma, and Pi Tau Sigma, and a junior member of the A.S.M.E.

WILLIAM R. BENNETT, B.S. in E.E., Oregon State College, 1925; M.A., Columbia University, 1928; Ph.D., Columbia, 1949. Bell Telephone Laboratories, 1925-. His early Laboratories projects included work on wire transmission problems, particularly the development of terminal apparatus in the voice and telegraph range, the design of circuits for television, and submarine cable telephony. Concerned with the coaxial cable in 1935, he spent several years working on the requirements and measuring techniques applicable to the load rating of multi-channel repeaters. His work during World War II was directed to a

number of military projects. Since then he has concentrated on pulse code modulation and general transmission problems. Member of the A.I.E.E., I.R.E., The American Physical Society, Tau Beta Pi, Eta Kappa Nu and Sigma Xi.

A. N. GRAY, Bell Telephone Laboratories, 1922-1929; Western Electric Company, 1930-. Mr. Gray, Assistant Superintendent, Development Engineering, Point Breeze since 1946, is engaged in the development of new equipment and processes. He was Manufacturing Engineer, Rubber Covered Wire, throughout the period of World War II when the Western was heavily loaded with the manufacture of communications items for the Armed Services. He is a member of the A.S.T.M., being Western's representative from Point Breeze, and is assigned to Committee D-11 on Rubber and Rubber Products.

L. N. HAMPTON, Cooper Institute of Technology; Experimental Department, Otis Elevator Company, Engineering Department; Western Electric Company and Bell Telephone Laboratories, 1917-. In the Western Electric Company's Apparatus Development Department, he designed Signal Corps apparatus for the detection of airplanes and submarines. Later, in Switching Apparatus Development, he was in charge of the development of apparatus for use in the telephone plant. After World War II and work on airborne radar and computing systems for military projects, he was engaged in the development of train-dispatching apparatus, cameras for photographing subscribers' message registers and the cam switching panels of the overseas radio privacy systems. He also was responsible for the development of the trouble recorder used in the 4 and 5 crossbar systems and the apparatus aspects of the card translator. More recently he has been active in the development of components for guided missiles. Member of the A.S.M.E. and the General Society of Mechanics and Tradesmen of New York; secretary, Foundation for Homeopathic Research.

H. R. HUNTLEY, B.S. in E.E., University of Wisconsin, 1921. Wisconsin Telephone Co. 1917-1930, except for a leave of absence to complete education begun earlier at Leland Stanford University and continued at the University of Wisconsin. Leaving The Wisconsin Telephone Company where he was Transmission Engineer, Mr. Huntley came to the Foreign Wire Relations Section of the Operating and Engineering Department of American Telephone and Telegraph Company in 1930. In 1942 he transferred to the Transmission Section and has been Transmission Engineer since 1951.

LUTHER W. HUSSEY, A.B., Dartmouth College, 1923; M.A., Harvard University, 1924. Union College, Instructor, 1924-30; Bell Telephone Laboratories, 1930-. Mr. Hussey was first engaged in research on non-linear resistive and reactive devices such as copper-oxide and germanium diodes. He worked on the development of a non-linear coil for the magnetic pulse generator and the harmonic generator in the megacycle range. He has been concerned with the development of modulating devices, negative impedance circuits, and switching and computer devices, and is currently associated with an electronic apparatus development group working on transistors and transistor circuits. He is a member of the I.R.E.

ROBERT L. KAYLOR, B.S. in E.E., University of Michigan, College of Engineering, 1927. Detroit Edison Company, 1922-27; American Telephone and Telegraph Company, Development and Research Department, 1927-34; Bell Telephone Laboratories, 1934-. At A. T. and T. he was engaged in field testing of new telephone apparatus and fundamental studies of noise and cross-induction in telephone circuits. He continued these and related studies after transferring to the Laboratories, his work including fundamental studies of methods of measuring radio noise. During World War II he was Signal Officer with several Army and Air Corps organizations, and is now a Lieutenant Colonel in the Air Force Reserve. Mr. Kaylor returned to the Laboratories in 1945 to do field trials of radio relay systems, and analysis and measurement studies in the 4,000-mc range. More recently he has been engaged in classified military projects. Member of the A.I.E.E., Associate of the I.R.E.

G. E. MURRAY, Western Electric Company, 1936-. Mr. Murray has been active in the development of equipment and processes for the electroforming project and is in charge of the electrochemical development group. During World War II, he was engaged in the manufacture of rubber covered wire and communications items for the Armed Services. He is a member of the American Chemical Society.

JAMES B. NEWSOM, Western Electric Company and Bell Telephone Laboratories, 1920-. After four years of military service in World War I, he joined Western Electric, directing his attention to the development of manual telephone systems and the panel telephone system. Since the incorporation of the Laboratories in 1925, he has been a member of what is now Switching Systems Development II and has devoted time

to the design of panel and crossbar systems, crossbar tandem and toll crossbar systems. During World War II he was a Lieutenant Commander in the U. S. Navy, assigned to the Naval Research Laboratory in Washington, D. C. Since 1946, Mr. Newsom has been in charge of a group concerned with the development of toll crossbar senders, decoders, translators and markers.

F. W. STUBNER, B.S., Cooper Union, 1930. Bell Telephone Laboratories, 1929-. Mr. Stubner joined the Laboratories' research drafting department and became a design engineer concerned with the design and building of apparatus and testing equipment for telephone instruments and submarine cable. Transferring to the Electronic Apparatus Development Department in 1940, he worked on the design of vacuum tubes, magnetic switches, and glasswork for the carbon deposited resistor. Since 1944 he has been associated with the applied mechanics laboratory, responsible for strength tests on vacuum tubes, shock and vibration studies, and associated design assignments. He transferred to Allentown, Pa., in 1948. Member of the Engineers Club of the Lehigh Valley and the Society for Experimental Stress Analysis.

A. S. WINDELER, B.S., Rutgers University, 1930; Bell Telephone Laboratories, 1930-. Mr. Windeler has been engaged in the design and development of toll cable, including coaxial, video pair, and microwave, types. He is currently in charge of a group concerned with the development of expanded polyethylene insulated conductors for multipair cable.

H E B E L L S Y S T E M
Technical Journal
DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXII

NOVEMBER 1953

NUMBER 6

Design Theory of Junction Transistors	J. M. EARLY 1271
Transistor Oscillator for Use in Multifrequency Pulsing Current Supply	F. E. BLOUNT 1313
Ferrites in Microwave Applications	J. H. ROWEN 1333
Cold Cathode Tubes for Transmission of Audio Frequency Signals	M. A. TOWNSEND AND W. A. DEPP 1371
Balanced Polar Mercury Contact Relay	J. T. L. BROWN AND C. E. POLLARD 1393
Dynamic Measurements on Electromagnetic Devices	M. A. LOGAN 1413
Selenium Rectifiers — Factors in Their Application	J. GRAMELS 1469
Arcing of Electrical Contacts in Telephone Switching Circuits Part II — Characteristics of the Short Arc	M. M. ATALLA 1493
Abstracts of Bell System Papers Not Published in this Journal	1507
Contributors to this Issue	1515

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

S. BRACKEN, *President, Western Electric Company*

F. R. KAPPEL, *Vice President, American Telephone
and Telegraph Company*

M. J. KELLY, *President, Bell Telephone Laboratories*

EDITORIAL COMMITTEE

E. I. GREEN, *Chairman*

A. J. BUSCH

F. R. LACK

W. H. DOHERTY

W. H. NUNN

G. D. EDWARDS

H. I. ROMNES

J. B. FISK

H. V. SCHMIDT

R. K. HONAMAN

G. N. THAYER

EDITORIAL STAFF

J. D. TEBO, *Editor*

M. E. STRIEBY, *Managing Editor*

R. L. SHEPHERD, *Production Editor*

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; Alexander L. Stott, Treasurer. Subscriptions are accepted at \$3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXII

NOVEMBER 1953

NUMBER 6

Copyright, 1953, American Telephone and Telegraph Company

Design Theory of Junction Transistors

By J. M. EARLY

(Manuscript received September 8, 1953)

The small signal ac transmission characteristics of junction transistors are derived from physical structure and bias conditions. Effects of minority carrier flow and of depletion layer capacitances are analyzed for a one dimensional model. The ohmic spreading resistance of the base region of a three dimensional model is then approximated. Short circuit admittances representing minority carrier flow, depletion layer capacitances, and ohmic base resistance elements are then combined into an equivalent circuit. Theoretical calculations are compared to observations for two typical designs.

1.0 INTRODUCTION

1.1 General

Junction transistors have been in commercial production for nearly a year. A detailed understanding of their behavior is necessary both for the increasingly exacting requirements of modern circuit engineering and for the wise design of improved types. Design theory, by relating function to structure, can serve both these needs.

The principal object of this paper is to develop in logical fashion a design theory for junction transistors. The product of the development is an equivalent circuit, founded on device physics, which predicts the circuit characteristics of junction devices in a simple and intelligible fashion. Although attention is concentrated on small signal transmission performance, some large signal aspects are also examined.

1.2 Method and Assumptions

The usefulness of the junction transistor derives primarily from the flow of holes or electrons across two closely-spaced p - n junctions, one of which is biased in the forward or conducting direction while the other is biased in the reverse or non-conducting direction. Development of design theory begins quite properly with analysis of this mechanism, which is considered, for simplicity, as a problem in the flow of holes and electrons in one dimension, at right angles to the p - n junctions. In the analysis, it is assumed that these carriers are controlled largely by the voltages applied to the junctions and that they move principally by diffusion. The dependence of the diffusion currents on the junction voltages is reduced to a set of two terminal-pair short-circuit admittances, which form the initial and most important segment of the equivalent circuit model for the junction transistor.

Practical transistors have not only the very useful transistoring mechanism mentioned above, but also passive capacitances across the charge depletion layers which separate the p and n regions at each junction. These capacitances limit the useful frequency range of transistors and must be considered in any practical theory. In the synthesis of the equivalent circuit, these capacitances are placed in parallel with the short-circuit input and output admittances which represent the flow of diffusing holes and electrons.

A further limitation on performance is imposed by the ohmic or body spreading resistance of the base region. The base current of the transistor, in flowing from the region between the emitter and collector to the base contact, develops a base contact to emitter voltage which seriously limits the frequency response. Calculation of these effects requires the assumption of flow paths for the base current. The circuit elements representing base spreading resistance effects appear in series in the base leg of the equivalent circuit.

1.3 Existing Design Theory

W. Shockley's classic paper* announcing the junction transistor also initiated the design theory. Diffusion effects for dc and low frequencies were analyzed, and formulae for depletion layer capacitances were developed. The mechanism of the frequency cutoff of the current transmission (α) was reported in a subsequent article†, and the effects of

* W. Shockley, The Theory of p - n Junctions in Semiconductors and p - n Junction Transistors, B.S.T.J., **28**, p. 435.

† W. Shockley, M. Sparks, and G. K. Teal, The p - n Junction Transistors, Phys. Rev., **83**, p. 151, July, 1951.

ohmic resistance of the base region were discussed briefly. Still more recently, the dependence of base thickness on collector voltage was used to explain output and feedback effects*. The present paper is both a consolidation and an extension of the earlier works and borrows freely from them. The diffusion current analysis of Appendix A is patterned after Shockley's.

1.4 Scope

The design theory developed here is not complete, even for small signal ac transmission. In particular, effects of large carrier emission densities are not considered, nor are the effects of non-parallel junction arrangements. Despite these omissions, it is hoped that the theory developed will be both useful and instructive to those engineers charged with transistor device and transistor circuit design.

2.0 METHODS AND ASSUMPTIONS

2.1 General

In developing the design theory, it is convenient to break the transistor down into several internal electronic functions and to consider their dependence on structure and materials individually. These functions are then fitted together and used to predict the terminal electrical characteristics. With this approach, it seems proper to describe separately the methods and assumptions used in analyzing each of the functions.

2.2 List of Symbols

The symbols listed here are used in the body of the paper. A separate list for Appendix A appears at the end of that section.

Emitter and collector currents are assumed to flow inward at the corresponding terminals, in accord with the convention usually used for transistors.

a = gradient of $(N_d - N_a)$, usually given in atoms/cm⁴.

a_{ce} = short-circuit forward current transfer constant for theoretical one-dimensional transistor.

C_c = collector to base capacitance with emitter open-circuit ac.

C_{se} , C_{sc} = hole storage or diffusion capacitances at emitter and collector. These capacitances are directly related to the current trans-

* J. M. Early, Effect of Space-Charge Layer Widening in Junction Transistors, I.R.E., Proc., 40, pp. 1401-1406, Nov., 1952.

mission cutoff frequency and may be used as an alternative characterization of that quantity.

C_{te} , C_{tc} = theoretical depletion layer capacitances of emitter and collector.

D , D_p , D_n = diffusion constants for minority carriers, usually given in cm^2/sec .

$f_\alpha = D/\pi w_0^2$ = current transmission or alpha cutoff frequency.

g_{ee} , g_{ce} , g_{ec} , g_{cc} = low-frequency conductance components of y 's given below.

h 's = set of two terminal-pair parameters, defined by Guillemin, Communication Networks, **2**, p. 137, John Wiley and Sons.

h_{11} = short circuit input impedance.

h_{21} = short circuit forward current transfer ratio.

h_{12} = open circuit feedback voltage ratio.

h_{22} = open circuit output admittance.

I_b = average or dc base current.

I_{pe} , I_{ne} , I_{pc} , I_{nc} = hole and electron components of average or dc emitter and collector currents.

I_{peo} = emitter reverse current when collector is also reverse biased.

J_e = emitter current density in amperes/ cm^2 .

k = Boltzmann's constant.

kT/q = average thermal energy per carrier, approximately 0.026 electron-volts at 25°C .

L , L_p , L_n = diffusion length or average distance a minority carrier will diffuse before recombining; average distance diffused in one lifetime (τ).

N_A , N_D = concentration of acceptor and donor atoms in semi-conductor, usually in atoms/ cm^3 .

n = concentration of electrons/ cm^3 .

n_i = electron concentration which would exist in the semi-conductor at thermal equilibrium if donor and acceptor concentrations were zero.

n_p = thermal equilibrium concentration of electrons in p -region.

p = hole concentration/ cm^3 .

p_i = hole concentration which would exist in the semi-conductor at thermal equilibrium if donor and acceptor concentrations were zero.

p_n = thermal equilibrium concentration of holes in n -region.

q = electronic charge, 1.6×10^{-19} coulombs.

q/kT = see kT/q .

r_b', r_{b1}', r_{b2}' = ohmic spreading resistances of base region, specifically, the effective base to emitter feedback resistances for diffusion currents and for collector capacitance currents.

r_1, r_2, r_3 = geometrical radii in transistor of Fig. 2(b).

T = temperature in $^{\circ}\text{K}$.

V_c = average or dc collector to base voltage.

V_c' = electrostatic potential across collector depletion region.

V_s = electrostatic potential across emitter depletion layer at thermal equilibrium (no biases applied).

v_c = small signal ac collector to base voltage.

w, w_0 = base region thickness.

w_1, w_2, w_3 = base region thicknesses in transistor of Fig. 2(b).

x_m = thickness of collector depletion region.

$y_{ce}, y_{ce}, y_{ec}, y_{cc}$ = theoretical short circuit input, forward transfer, feedback, and output admittances for one-dimensional transistor.

α, α_0 = short-circuit emitter to collector current transfer ratio and its low-frequency value.

α^*, α_0^* = collector junction current multiplication ratio and its low-frequency value.

β, β_0 = current transport ratio across base region and its low-frequency value.

γ, γ_0 = current emission ratio at emitter and its low-frequency value.

ϵ_0 = dielectric constant of vacuum, 8.854×10^{-14} farad/cm.

k = relative dielectric constant, ϵ/ϵ_0 .

ρ, ρ_b = resistivity, base region resistivity.

σ_{nc}, σ_{pc} = conductivities produced by electrons and holes in collector region.

τ, τ_n, τ_p = lifetimes of minority carriers.

μ_{bc} = constant of feedback generator used to characterize modulation of dc base spreading resistance.

$\omega = 2\pi f$ = angular frequency in radians.

$\omega_\alpha = 2\pi f_\alpha = 2D/w_0^2$ = alpha or current transmission cutoff frequency in radians.

2.3 Minority Carrier Admittances

An admittance representation of minority carrier diffusion is a way of writing the dependence of the diffusion currents on the junction potentials. To obtain this dependence analytically, the minority carrier densities on both sides of each of the two depletion layers (emitter and collector) are assumed to be exponential functions of the junction voltages. This exponential dependence is a result of the normal thermal distribution of hole and electron energies. The carrier diffusion currents are computed directly from the gradients of the minority carrier densities at the depletion layer surfaces. Since the gradients of the carrier densities are affected by many conditions besides the junction voltages, additional assumptions are necessary. Their nature and pertinence may be seen from consideration of the normal operation of a junction transistor.

The three principal regions of a junction transistor, the emitter, the base or control, and the collector, are indicated in Fig. 1. These regions are separated by transition regions in which the conductivity type changes either gradually or abruptly from *p*-type to *n*-type. Roughly coincident with these transition regions are the emitter and collector depletion layers across which the emitter and collector voltages appear

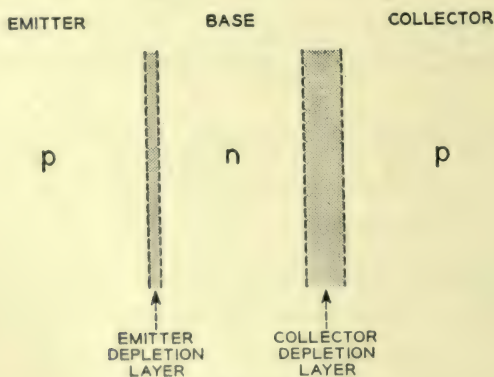


Fig. 1 — *p-n-p* transistor.

when the unit is biased. In normal operation for the p - n - p transistor shown, the emitter is biased positive with respect to the base so that a current of holes is injected into the base from the emitter. The collector is biased negative with respect to the base so that the holes diffusing across the base from the emitter are collected whenever they reach the edge of the collector depletion layer.

In the analysis each of the three major regions is assumed to have a uniform resistivity, ρ ; a diffusion constant for minority carriers, D , which is a measure of the speed with which injected carriers will diffuse; and a lifetime for minority carriers, τ . This lifetime is the average time which a minority carrier remains free before recombining with a majority carrier. The minority carrier density in each region is assumed to have a thermal equilibrium value in the absence of applied potentials. The density is increased or decreased exponentially from this value by the applied potentials. The base layer is assumed to have a thickness, w , which is dependent on the collector voltage V_c . An increase of collector voltage increases the collector depletion region thickness, x_m , thus decreasing the base thickness. The rate at which base thickness changes with collector voltage is determined by the nature of the transition from base to collector. For gradual transitions, the rate of the transition is important, while for abrupt or step transitions the rate of change of base thickness with collector voltage is determined by the base region and collector region resistivities.

Determination of the diffusion currents from minority carrier density gradients requires determination of minority carrier densities everywhere in the three principal regions of Fig. 1. These are obtained by solving a continuity equation for carrier flow in each region, subject to the applied junction potentials and other assumptions described above. It must be pointed out that, in normal operation, there may be a significant flow of electrons to the emitter and from the collector in the p - n - p transistor of Fig. 1. Small signal ac diffusion currents are determined by assuming small signal variations of the junction voltages and discarding all but first-order ac terms from the diffusion currents.

Results of the analysis are given in Section 3.0, and the analysis appears in Appendix A.

2.4 Depletion Layer Capacitances

In a p - n junction with no bias potential applied, there is a tendency for holes to diffuse into the n -region and for electrons to diffuse into the p -region. This creates a slight unbalance of charge in the two regions

and the resulting electrostatic potential keeps each type of carrier in its own region. The potential appears across a thin layer separating the two regions. In this depletion layer, the hole density is lower than in the p -region and the electron density is lower than in the n -region, and there is a net charge density. Acceptor and donor atoms are not neutralized by mobile charge as they are in the p - and n -regions, but instead serve to terminate the field of the electrostatic potential. Application of external potential across the junction changes the electrostatic potential, and by exposing more or fewer fixed (donor and acceptor in equal number) charges widens or narrows the depletion layer.

The passive capacitance of this region is simply that of a parallel plate condenser having a plate spacing equal to the layer thickness. Calculation of this capacitance is explained in Section 3.0, following the discussion of the minority carrier diffusion admittances.

2.5 Base Spreading Resistance*

Implicit in the one-dimensional analyses described above is the assumption that the base region is everywhere at the same potential. Actually, since the emitter and collector currents are not equal, current must flow through the base region parallel to the junctions. Because the base region has finite, rather than zero resistivity, this current produces transverse voltage drop in the base region.

It is assumed that the most important effect of these voltage drops is the feedback produced between the base contact and the emitter junction. In consequence, each of the ohmic base resistances studied is defined as the quotient of an average voltage between base contact and emitter junction divided by the current producing it. The need for defining more than one feedback base spreading resistance results from the fact that the base current has two principal ac components. One of these is the difference between the emitter and collector minority carrier diffusion currents. The other is the collector depletion layer capacitance current. The feedback effects of these two currents on the emitter junction are the same only when the flow paths of the two currents through the base region are the same. Consequently, the representation of base resistance effects is somewhat more complicated in transistors where the flow paths differ than in those where they are identical or nearly so.

* The majority carrier resistance of the base region for base current flow parallel to the junctions. The word "spreading" was suggested by the base contact geometry of Fig. 2(a) and readily distinguishes this resistance from the "base resistance" of the familiar tee network, which was long believed to be identical with it. It is not. See Sections 5.2 and 5.3.

Fig. 2(a) shows a structure for which the flow paths to the base contact are substantially the same for all components of the base current. Both the collector capacitance current and the diffusion loss base current enter the base region substantially uniformly over the entire area and follow the same path to the base contact. In Fig. 2(b) these two currents have quite different flowpaths and the associated feedback resistances are likewise very different. The general method of calculation is, however, the same in both cases.

Another important effect is associated with modulation of the dc voltage drop in the base region. The base current ordinarily has a dc as well as an ac component, and a dc voltage drop occurs between the base contact and the emitter junction. Since the base region thickness changes when collector voltage changes, the dc resistance of the base region is modulated by the collector voltage, producing a modulation of the dc voltage between base contact and emitter.* This effect is most easily represented by an ac voltage generator in series with the base

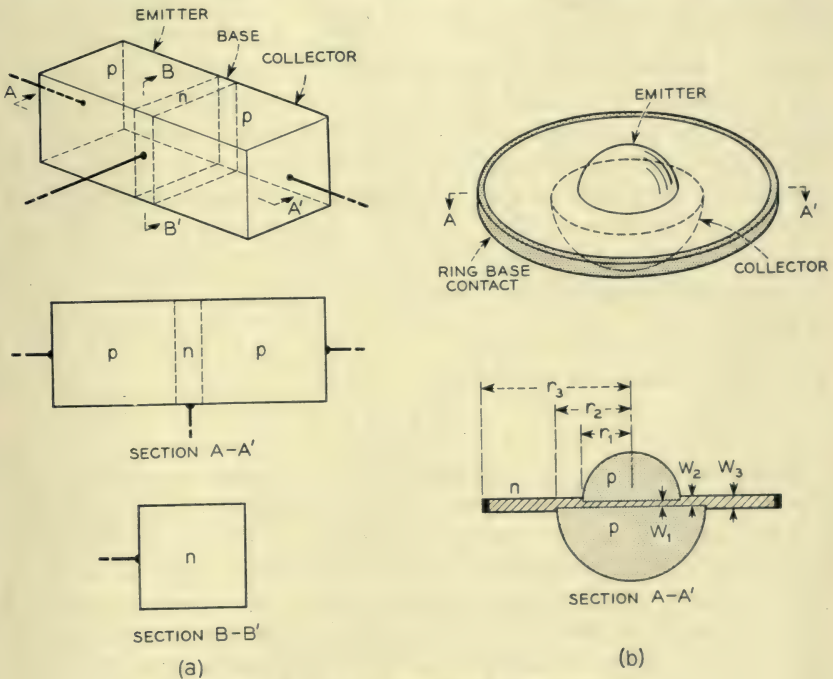


Fig. 2 — *p-n-p* transistor structures.

* This effect was first pointed out by J. N. Shive.

contact. The voltage is computed as the product of the dc base current and the modulation of the dc base resistance. This modulation can be calculated from the base region resistivity and the dependence of base thickness on collector potential.

2.5 *Summary of Methods*

In developing the design theory, simple physical assumptions are made concerning the behavior of the charge carriers in the semiconductor. The transistor is studied as a one-dimensional problem and the per unit area electrical characteristics of the one dimensional structure are computed. The effects of current flow within the base region parallel to the junctions are then calculated for a three-dimensional model. Finally, the equivalent circuit representations of these electronic functions are combined in structural fashion to give the terminal electrical characteristics of the junction transistor triode.

It should be noted that the base region thickness between emitter and collector is assumed uniform, and that design theory has not been extended here to cover the case of non-uniform thickness. Likewise, edge effects at the emitter and surface effects in general are neglected. These omissions were made for mathematical simplicity and are necessary omissions in a one-dimensional analysis. The place of surface leakage among the electronic functions is discussed at the end of Section 4.0. Analysis of the effects of sharp discontinuities in base layer thickness requires new solutions to the continuity equation but gradual changes in thickness can be accounted for by averaging over the active area of the transistor the short circuit admittances which are the subject of the next section.

3.0 ONE-DIMENSIONAL TRANSISTOR

3.1 *General*

This section deals with the small signal transmission electronics of the structure of Fig. 1. It is assumed that the emitter is biased to provide a flow of carriers into the base and that the collector is reverse biased sufficiently so that no majority carriers can diffuse out of the collector region into the base region (a reverse voltage of 0.5 volts is more than enough to prevent this). The four admittances associated with minority carrier flow and the two depletion layer capacitances are indicated in Fig. 3. In each case, the design expressions are given first in their most exact form and are progressively simplified. For convenience in discussion and comparison, current densities per unit area rather than

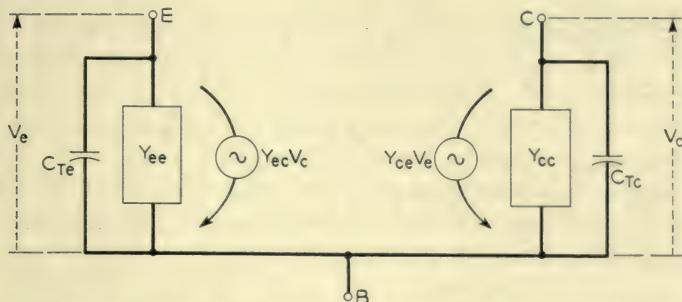


Fig. 3 — Theoretical equivalent circuit for "one-dimensional" transistor.

currents are used and the admittances and barrier capacitances are written on a per square centimeter basis. It will be noted that transverse voltage drops resulting from base spreading resistance are ignored in developing the expressions of this section.

A number of physical mechanisms are involved in the admittances for minority carrier flow. The forward current across the emitter junction rises as an exponential of the emitter to base region voltage. This is the result of the thermal energy distribution of minority carriers and is common to all thermionic emission. A natural effect of this exponential dependence is that a given change in the voltage results in a fixed per cent change in the current, thus producing an ac admittance which is proportional to the average or dc emitter current. For several reasons not all of the current which flows through the emitter junction is collected. First, some of the emitter current consists of electrons diffusing into the emitter body and of displacement current through the emitter depletion layer capacitance. These effects are expressed in the emission factor or γ , which is the ratio of the hole current injected into the base region to the total emitter current. Further, some of the injected holes recombine in the base layer. Those which are collected suffer a transit delay which results in a phase difference between emitter and collector currents. These two effects are summed up in the forward current transport factor or β , which is the ratio of the minority carrier current reaching the collector to that injected by the emitter. Finally, the current of holes entering the collector from the base gives rise to a much smaller flow of electrons from the collector to the base, thus producing a collector multiplication factor or α^* , which is the ratio of total carrier current crossing the collector junction to the hole current entering it from the base.

Although the most important minority carrier flow originates at the emitter, this flow is altered by changes in collector reverse voltage. As the collector reverse potential is increased, the base region becomes

narrower because of widening of the collector depletion layer. This reduction in base thickness permits more emitter current to flow for a fixed emitter to base voltage, so that both emitter and collector currents are increased, thus producing output and feedback admittances. However, since large changes in collector potential are required to produce small changes in base thickness, relatively small changes in the junction currents are produced and the admittances are far smaller than those associated with change of emitter potential.

3.2 Diffusion Current Admittances

The short circuit two terminal-pair admittances associated with the diffusion of minority carriers in the structure of Fig. 1 are

$$y_{ee} = \frac{q}{kT} \left[(I_{pe} - I_{pe0}) \frac{(1 + i\omega\tau_p)^{1/2} \tanh w_0/(D_p\tau_p)^{1/2}}{\tanh [(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}]} \right. \\ \left. + I_{ne} (1 + i\omega\tau_{ne})^{1/2} \right]$$

$$y_{ce} = -\frac{q}{kT} \left[(I_{pe} - I_{pe0}) \left[\frac{(1 + i\omega\tau_p)^{1/2} \tanh w_0/(D_p\tau_p)^{1/2}}{\sinh [(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}]} \right] \left[1 + \frac{\sigma_{nc}}{\sigma_{pc}} \right] \right]$$

$$y_{ec} = -\frac{\partial w}{\partial V_c} \frac{(1 + i\omega\tau_p)^{1/2}}{(D_p\tau_p)^{1/2} \sinh [(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}]} I_{pc}$$

$$y_{cc} = -\frac{\partial w}{\partial V_c} \frac{(1 + i\omega\tau_p)^{1/2}}{(D_p\tau_p)^{1/2} \tanh [(1 + i\omega\tau_p)^{1/2}w_0/(D_p\tau_p)^{1/2}]} I_{pc} \left[1 + \frac{\sigma_{nc}}{\sigma_{pc}} \right]$$

in which I_{pe} , I_{ne} , and I_{pc} are average hole and electron currents at the emitter and collector junctions and I_{pe0} is the emitter reverse current measured with both junctions reverse biased. w_0 is the (time) average base region thickness and V_c is the average collector voltage. The depletion layer capacitances which shunt the input and output diffusion admittances, y_{ee} and y_{cc} , are discussed in Section 3.3.

Some general features of the admittances may be noted at once. The expressions are similar to those for a section of lossy transmission line, but differ significantly in the ordering of the magnitudes of the terms. Each of the admittances is proportional to a dc current, and in fact, since I_{pc} is ordinarily ninety per cent or more of I_{pe} , approximately the same dc current. This effect, which results from the exponential dependence of emitter current on emitter potential, can be related to the transmission line analogy by the argument that the increase in minority carrier density in the base region which accompanies increase of dc current lowers the characteristic impedance of the transmission line.

Finally, the last two terms, the feedback and output admittances, are smaller than the first two by approximately

$$\frac{kT}{qw_0} \frac{\partial w}{\partial V_c}$$

which is usually 10^{-3} or less, but the four form a rather symmetrical set.

The symmetry of the terms can be seen by removing the dissymmetries. The term $I_{ne} (1 + i\omega\tau_{ne})$ in y_{ee} results from diffusion of electrons into the emitter from the base and is the emission loss term which makes γ less than unity. The factor $(1 + \sigma_{nc}/\sigma_{pc})$ which appears in y_{ce} and y_{cc} is the collector multiplication factor α^* which results from the flow of electrons out of the collector body. If γ and α^* are assumed to be unity, the admittances assume a more symmetrical form. In that case $y_{ce}/y_{ee} = y_{ec}/y_{cc} = -\beta$, the base transport factor, and the forward and reverse current transmission ratios are identical. This can be seen more clearly if the hyperbolic expressions are replaced by the first two terms of their polynomial expansions, yielding

$$y_{ee} = g_{ee} \frac{(1 + i\omega/\omega_\alpha)}{(1 + i\omega/3\omega_\alpha)}$$

$$y_{ce} = \frac{-\beta_0 g_{ee}}{1 + i\omega/3\omega_\alpha}$$

$$y_{ec} = \frac{-\beta_0 g_{cc}}{1 + i\omega/3\omega_\alpha}$$

$$y_{cc} = g_{cc} \frac{(1 + i\omega/\omega_\alpha)}{1 + i\omega/3\omega_\alpha}$$

where

$$g_{ee} = qI_e/kT, \quad g_{cc} = -\frac{1}{w} \frac{\partial w}{\partial V_c} I_{pc} \left(1 + \frac{1}{2} \frac{w_0^2}{D_p \tau_p} \right),$$

$$\omega_\alpha = \frac{2D_p}{w_0^2} \quad \text{and} \quad \beta_0 = 1 - \frac{1}{2} \frac{w_0^2}{D_p \tau_p}$$

is the low frequency value of the base transport factor. A lumped parameter equivalent circuit for these simplified admittances is shown in detail in Fig. 4(a).

It is apparent here that a single parameter, $\omega_\alpha = 2D_p/w_0^2$, specifies the frequency variation of all four admittances. This frequency dependence, which has been commonly measured as the alpha cutoff frequency

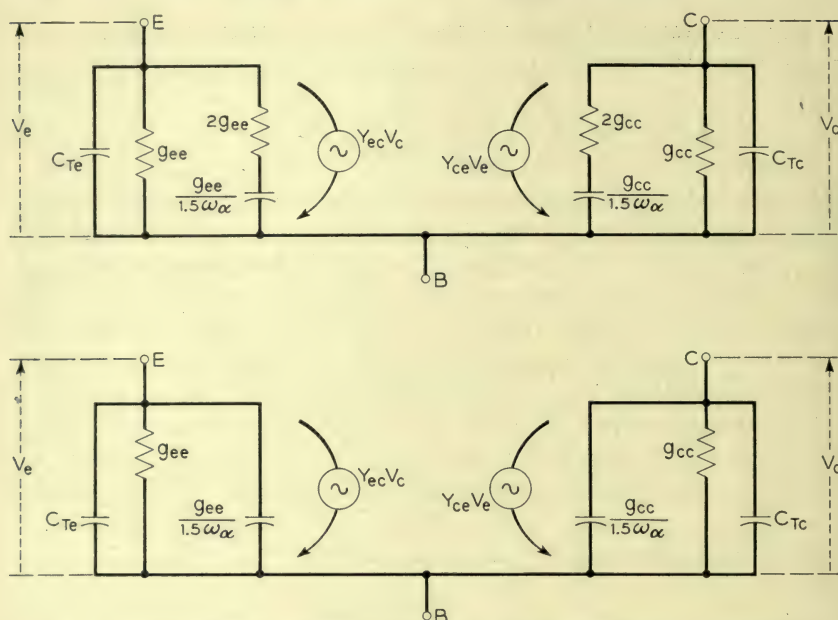


Fig. 4 — Simplified equivalent circuits for "one-dimensional" transistor.

or three db point† of the current transmission amplitude, appears in the admittances in the form of storage capacitance or as transfer delay. These result from minority carrier storage in the base region and are of the same nature as hole storage capacitances in semi-conductor diodes.

If desired, further simplification can be obtained by eliminating the effective power loss in the storage capacitances as shown in Fig. 4(b). In this arrangement, as in the more rigorous forms preceding it, the forward transfer admittance y_{ce} is a little smaller than the input admittance y_{ee} and the feedback admittance y_{ec} is not quite as large as the output admittance y_{cc} . This makes the circuit delta ($y_{ee}y_{oc} - y_{ce}y_{ec}$) greater than zero so that the structure is inherently stable when γ and α^* are unity.

3.3 Depletion Layer Capacitances

Even in the absence of applied potentials, there exists at useful p - n junctions a depletion layer in which the density of mobile carriers is low,

† R. L. Pritchard has pointed out that this three db point is 22 per cent larger than ω_α as defined here. See R. L. Pritchard, Frequency Variations of Current-Amplification Factor for Junction Transistors, Proc. I.R.E., **40**, p. 1476, Nov., 1952.

and across which there is a small electrostatic potential. The potential difference across the region exists primarily because of the difference in energy levels between the conduction band in which the mobile electrons exist and the valence band in which the mobile holes move. The potential difference depends on the densities of holes and electrons in the p and n regions, but it cannot exceed the energy gap or difference in band levels, so long as the junction is at equilibrium.

If a reverse voltage is applied to a junction, the applied voltage appears principally across the depletion region, where it strengthens the electric field by widening the barrier region so as to bring more fixed donor and acceptor charges into the field. It is obvious that the depletion region has a capacitance, since an electric field exists across it. This capacitance decreases with increase of reverse voltage, since the capacitance is charged not by bringing mobile carriers to fixed electrodes but rather by widening the region to include new fixed charges from the semiconductor regions on each side.

Both emitter and collector capacitances may be calculated easily for the principal cases of published engineering interest, the graded transition and the abrupt or step transition. For graded transitions, the depletion layer is in a region of linearly changing fixed charge density (zero at the center of the layer). The step junction has the barrier layer almost entirely in either a p or an n region of uniform fixed charged density since the fixed charged density in the other region is usually so large that the field effectively terminates at its surface.

The general expression for barrier capacitance is

$$C = \frac{\kappa\epsilon_0}{x_m}$$

where x_m is barrier thickness and the other symbols have conventional meanings. In the case of graded junctions, this becomes

$$C = \frac{\kappa\epsilon_0}{2} \left(\frac{2qa}{3\kappa\epsilon_0 V_c'} \right)^{1/3}$$

where a is the rate of change of fixed charge density in charges per cm^3 per cm. For step junctions, the relation is

$$C = \kappa\epsilon_0 \left(\frac{q(N_d - N_a)}{2\kappa\epsilon_0 V_c'} \right)^{1/2}$$

where $(N_d - N_a)$ is the fixed charge density in the low charge density region, ordinarily the base region in transistors. The potential V_c is the electrostatic potential across the depletion layer and under bias conditions is

the sum of the equilibrium barrier potential and whatever part of an applied external potential is not lost in IR drops in the p and n regions. For reverse bias conditions, the sum of the equilibrium potential and the entire applied potential is almost always a satisfactory approximation, but for equilibrium and forward bias conditions, the barrier potential may be computed more easily from the minority carrier densities immediately adjacent to the depletion layer.

The equilibrium barrier potential may be calculated by

$$V_s \simeq \frac{kT}{q} \ln \frac{n_i}{n_p} \cdot \frac{p_i}{p_n}$$

in which n_n , p_p and n_i , p_i are, respectively, equilibrium and intrinsic carrier densities in the two regions. The total depletion region potential for a forward biased emitter junction may be calculated by the same expression but the hole density in the base region side of the emitter depletion layer is not the equilibrium value p_n but rather

$$p = \frac{w_0 J_e}{q D_p}$$

which is simply a modified form of the equation for diffusion current. Because of the large difference in the voltages across the barriers, emitter depletion region capacitance per unit area is ordinarily 2–20 times larger than the collector depletion region capacitance per unit area. It should not be assumed that emitter depletion region capacitance is the more important. It contributes but a small fraction of the emitter admittance, while the collector depletion layer capacitance contributes greatly to collector admittance.

3.4 Summary

A one-dimensional study of the small signal transmission properties of the junction transistor shows two terminal-pair short circuit admittances which are closely proportional to the dc bias currents. The input and output admittances of this set are shunted by depletion layer capacitances which are essentially passive circuit elements. A major feature of the diffusion current admittances is the presence of a single frequency determining factor, the alpha cutoff frequency. Comparison of the diffusion or storage capacitances with the depletion region capacitances shows that the collector depletion region capacitance is usually much larger than its storage capacitance, while at the emitter the reverse is true. This results from the fact that the output and feedback minority

carrier admittances are many orders of magnitude smaller than the input and forward transfer terms. The output and feedback admittances decrease with increase of collector reverse voltage in much the same way that collector capacitance decreases, while the input and forward transfer admittances are but little affected.

This simple picture of the junction transistor is incomplete in that it ignores the important effects of majority carrier resistance in the base region. These effects are discussed in the next section.

4.0 EFFECTS OF BASE REGION RESISTANCE

4.1 *General*

The resistance of the base region to the flow of currents parallel to the emitter and collector junctions is important primarily because of voltage developed between the base contact and the emitter junction. Necessarily associated with this is power dissipation which is usually, however, of less importance than the feedback effect of the voltage.

The primary factors in determination of base region feedback voltages are the materials and geometry of the base region and the flow paths for the transverse currents moving through the base region to the base contact. Reduction of the base region resistance effects to equivalent circuit elements permits them to be incorporated in the equivalent circuit obtained in the previous section. This circuit is then an essentially complete model for the electronic mechanisms or functions in junction transistor triodes.

4.2 *Base Region Currents*

Two of the three principal components of the base current have nearly identical origin and flow paths, while the third component differs in origin and may differ greatly in flow paths to the base contact. The dc component of the base current arises principally from recombination of injected holes in the portion of the base between emitter and collector,* as does also the ac component associated with the diffusion admittances. The ac component of base current required to charge the collector barrier capacitance, however, is introduced into the base uniformly over the surface of the collector. The ac component required to charge the emitter capacitance is both small and similar in flow paths to the first components discussed.

* In some transistors, much of the recombination occurs on the exposed surface of the base region. This surface recombination may be replaced by a reduction of volume lifetime for a one-dimensional analysis. This is not exact, but is a fair approximation.

For the structure of Fig. 2(a), all current components are introduced into the base layer relatively uniformly over its area. The dc component and the $(1 - \alpha)$ or diffusion component originate within the layer volume, while the collector capacitance current is introduced from the collector side of the layer, but this difference is of little consequence. The transverse current density is largest at the base contact and diminishes monotonically to zero at the layer edge opposite the contact. The feedback voltage to the emitter junction consequently rises most rapidly close to the contact and becomes nearly constant as the opposite side of the base region is reached.

For the structure of Fig. 2(b), significant differences in flow paths are apparent. The transverse current densities are zero at the center of the unit and rise out to the edge of the emitter. Here the dc and $(1 - \alpha)$ components of current density begin to decrease because of the increasing cross-section of the base region, while the collector capacitance components continues to increase because of the additional contribution from the larger collector area. All components decrease in density beyond the collector circumference. It is apparent that two feedback resistances are required to describe the separate electronic functions in this case.

4.3 Effective Feedback Resistances

The resistance effects of importance are not the series resistances to the flow of the various current components through the base region to the base contact, but rather the feedback resistances defined by the quotient of the average base contact to emitter junction voltage by the current component producing the voltage. The calculations necessary to determine such resistances are detailed in Appendix II.

For the structure of Fig. 2(a), a single equivalent resistance given by

$$r'_b = f \left(\frac{\rho_b}{w} \right) \simeq \frac{\rho_b}{w}$$

is sufficient. Both $(1 - \alpha)$ and C_c current components pass through this resistance en route to the base terminal. In the structure of Fig. 2(b), the $(1 - \alpha)$ or diffusion current component has an effective feedback resistance of†

$$r'_{b1} \simeq \rho_b \left[\frac{1}{8\pi w_1} + \frac{1}{2\pi w_2} \ln(r_2/r_1) + \frac{1}{2\pi w_3} \ln(r_3/r_2) \right]$$

while the collector capacitance component of base current has associated

† This value of r'_{b1} is an upper bound, since the assumptions are pessimistic.

with it a resistance

$$r'_{b_2} = \rho_b \left(\frac{1}{2\pi w_3} \ln(r_3/r_2) + (r_1/r_2)^2 \left[\frac{1}{2\pi w_2} \ln \frac{r_2}{r_1} + \frac{1}{8\pi} w_1 \right] \right. \\ \left. + \left[1 - \left(\frac{r_1}{r_2} \right)^2 \right] \frac{1}{4\pi w_2} \right)$$

For the proportions shown, the ratio of r'_{b_1} to r'_{b_2} is approximately 1.5 to one. Reasonable variations in the proportions can make this ratio as small as unity or as large as five.

A common feature in these resistances is the presence of the base region thickness w in the denominator. Since low values of base spreading resistance are desirable, this requirement opposes directly the primary requirement for a high alpha cutoff frequency — a thin base layer.

4.4 Base Resistance Modulation Feedback

As was mentioned previously, the dc voltage drop between the base contact and the emitter junction is modulated by the widening and narrowing of the collector depletion layer which alternately increases and decreases the dc base resistance. The resulting ac voltage may be represented by a voltage generator in series with the base resistance, as

$$\mu_{bc} v_c = I_b \frac{\partial r'_b}{\partial w} \frac{\partial w}{\partial V_c} v_c$$

where I_b is the dc base current, V_c is the dc collector voltage, v_c is the ac collector voltage, and $\mu_{bc} v_c$ is the feedback voltage.

For the structure of Fig. 2(a), the expression given earlier for r'_b may be differentiated to obtain the $\partial r'_b / \partial w$. The calculation of $\partial w / \partial V_c$ has been indicated in Section 3.0. The resulting expression for μ_{bc} is approximately

$$\mu_{bc} \simeq I_b \frac{r'_b}{w} \frac{\partial w}{\partial V_c}$$

Determination of this feedback effect for the structure of Fig. 2(b) is more difficult. Only those portions of the base region adjacent to the collector barrier are affected by the barrier widening, so that only the first two terms in the expression for r'_b , given above enter into the calculation. The value of μ_{bc} is

$$\mu_{bc} \simeq I_b \rho_b \left(\frac{1}{8\pi w_1^2} + \frac{1}{2\pi w_2^2} \ln \frac{r_2}{r_1} \right) \frac{\partial w}{\partial V_c}$$

This feedback term cannot be reduced to an expression containing r'_b ,

in the same simple way possible for the structure of Fig. 2(a), since the base resistance between the collector circumference and the base contact is not changed by collector depletion region thickness changes.

The details of the base contact may change the magnitude of this feedback greatly. In particular, poor placement of a bonded base contact may sometimes result in very large values of μ_{bc} . A point of general interest is that the sign of the feedback is determined directly by the direction of base current flow, while the phase is otherwise exactly that of the collector voltage.

4.5 Equivalent Circuit with Base Resistance Effects

The base resistance effects discussed above can be combined with the admittance circuits obtained in the one-dimensional study of Section 3.0 to give equivalent circuits for the three-dimensional structures of Fig. 2. The resulting representations, shown in Fig. 5, reflect the geometry of these structures since the base resistance elements are shown in a branch through which the base current must flow to reach the base terminal. All of the effects can be seen to give feedback, which is to say, coupling from the output circuit of the collector to the input loop of the emitter. This leads to complication of the electrical characteristics of the transistor, whose characteristics would otherwise be given by the electronic functions developed in the previous section.

Fig. 5(a) is an equivalent circuit which represents the small signal transmission properties of the structure of Fig. 2(a). It should be noted that a single base resistance is needed, through which all of the base current passes. The effects of surface leakage are indicated by the admittance Y_l from collector terminal to base resistance. The indicated uncertainty of placement of this leakage effect with respect to r_b' reflects the fact that the feedback resulting from the leakage depends on the position of the leakage with respect to the base contact.

The structure of Fig. 2(b) is represented by the circuit of Fig. 5(b). The separate base resistances r_{b_1}' and r_{b_2}' mirror the physical fact that the collector region is, on the average, closer to the base contact than is the emitter region. Leakage effects may be added to this circuit in the same manner as before.

Although the circuits of Fig. 5 represent primarily small signal ac transmission functions in the transistor, very similar representations may be employed for large signal functions. The principal changes occur in the diffusion admittances, which should be replaced by the diffusion currents written as functions of the barrier voltages, and in the capaci-

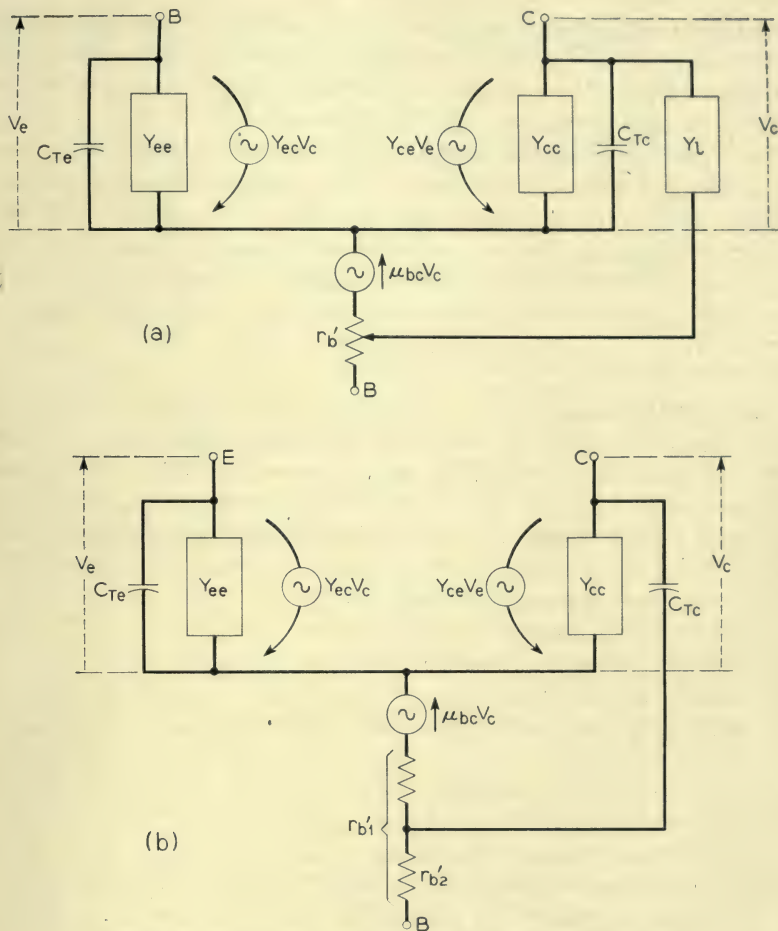


Fig. 5 — Equivalent circuits for junction transistors.

tances, whose variation with barrier potential must now be considered explicitly.

5.0 DESIGN CALCULATIONS

5.1 General

The understanding of junction transistors is now sufficiently advanced that this type of device is designable — the engineer may work from requirements to choice of structure and material. The limitations on designability are now primarily lack of adequate control over manufacturing variables together with the inherent characteristics of the

semi-conductor materials available to the designer. The designability of the major small signal transmission parameters of junction devices is illustrated below by a comparison of calculated and measured characteristics of both grown and fused junction devices.

The characteristics calculated are the theoretical short circuit conductances and current transmission ratios (alphas), the current transmission three db cutoff frequencies, the collector capacitances, the base region spreading resistances, and the low-frequency values of the "h" parameters. The latter are computed because of their ease of measurement and interpretation and are compared with measured values.

5.2 *n-p-n Drawn Junction Transistor*

The properties assigned to this unit are believed to be reasonable, but a close check on many of them is very difficult. Base region thickness was chosen as a reasonable value for current practice. The physical structure is that of Fig. 2(a), with the bar cross-section a square 20 mils on a side. It should be noted that the assumptions made are consistent; e.g., the grading of the collector junction is consistent with the collector and base region resistivities.

The physical parameters of interest are:

V_c Collector voltage, 4.5 volts

I_e Emitter current, -1 ma

w Base thickness (not including depletion layers), 3.6×10^{-3} cm or 1.41 mil

ρ_b Base resistivity, 1 ohm-cm

τ_b Electron lifetime in base, 10μ sec

ρ_e Emitter resistivity, 0.01 ohm-cm

τ_e Hole lifetime in emitter, 0.45μ sec

ρ_c Collector resistivity, 1.7 ohm-cm

τ_c Hole lifetime in collector, 30μ sec

a Concentration gradient in junctions, 3×10^{18} atoms/cm⁴

A Junction areas, 0.0025 cm²

The parameters of this unit are then

$$g_{es} = \frac{q}{kT} I_e = \frac{10^{-3}}{0.026} = 0.039 \text{ mhos}$$

$$\gamma_0 \simeq \frac{2}{1 + \frac{\rho_e}{\rho_b} w / \sqrt{D_p \tau_e}} = \frac{1}{1 + 0.008} = 0.992$$

$$\beta_0 \simeq \left[1 - \frac{1}{2} \frac{w^2}{D_n \tau_b} \right] = 1 - 0.0072 \simeq 0.993$$

$$\text{at } 25^\circ\text{C } \alpha_0^* \simeq \left[1 + \frac{1}{2} \frac{\sigma_{pc}}{\sigma_{nc}} \right] = 1 + 0.0003 \simeq 1.0003$$

$$70^\circ\text{C } \alpha_0^* \simeq 1.03$$

$$25^\circ\text{C } a_{ce} = \gamma_0 \beta_0 \alpha_0^* = 0.992 \times 0.993 \times 1.0003 = 0.985$$

$$70^\circ\text{C } a_{ce} = 1.015$$

$$\text{Now } \frac{\partial w}{\partial V_c} = \frac{x_m}{6V_c} = \frac{6 \times 10^{-4}}{6 \times 6} = 1.6 \times 10^{-5} \text{ cm/volt}$$

$$g_{ce} = -\frac{1}{w} \frac{\partial w}{\partial V_c} I_c \cosh (w/L_n)$$

$$\simeq \frac{kT}{qw} \frac{\partial w}{\partial V_c} g_{ee} = 0.026 \times 1.6 \times 10^{-5} \times 0.039$$

$$= 4.5 \mu\text{mhos}$$

The parameters which determine frequency response are

$$f_\alpha = \frac{\omega_\alpha}{2\pi} = \frac{D_n}{\pi w^2} = \frac{90}{\pi(3.6 \times 10^{-3})^2} = 2.45 \text{ meps}$$

$$C_T \simeq 2400 \mu\mu\text{fd/cm}^2 \text{ for collector}$$

$$C_T \simeq 6500 \mu\mu\text{fd/cm}^2 \text{ for emitter}$$

so that

$$C_{T_e} = 6 \mu\mu\text{fd}$$

$$C_{T_c} = 16 \mu\mu\text{fd}$$

The electron storage capacitances of the emitter and collector diffusion admittances of this n - p - n transistor are:

$$C_{se} = \frac{g_{ee}}{1.5\omega_\alpha} = \frac{0.039}{1.5 \times 2.45 \times 2\pi \times 10^6} = 1690 \mu\mu\text{fd}$$

$$C_{sc} = \frac{g_{cc}}{1.5\omega_\alpha} = 0.2 \mu\mu\text{fd}$$

It is obvious that diffusion effects are most important at the emitter, while depletion layer capacitance is important at the collector.

The ohmic base spreading resistance of this unit is given (probably within a factor of two) by

$$r'_b \simeq \frac{\rho b}{w} = \frac{1.0}{3.6 \times 10^{-3}} = 278 \text{ ohms}$$

The low-frequencies values of the hybrid ("h") parameters may be calculated from the conductances, etc., given above, but more direct methods are available and are used. The expressions required for both types of computation are given, however.

$$h_{21} \simeq -a_{ce} \simeq -0.985$$

$$h_{12} \simeq (g_{cc}/g_{ee}) + h_{22}r'_b + \mu_{bc}$$

$$\simeq \frac{kT}{qw} \frac{\partial w}{\partial V_c} \simeq 1.15 \times 10^{-4}$$

$$h_{22} \simeq \frac{g_{cc}g_{ee} - g_{ec}g_{ce}}{g_{ee}} = g_{cc}[1 - a_{ec}a_{ce}]$$

$$g'_c = h_{22} = \frac{I_e}{w} [2(1 - \beta_0) + (1 - \gamma_0)] \frac{\partial w}{\partial V_c} = 0.11 \mu mho$$

$$h_{11} \simeq \frac{1}{g_{ee}} + (1 - a_{ce})r'_b$$

$$\simeq 26 + (0.015)278 \simeq 30.5 \text{ ohms}$$

These parameter values are in the range commonly encountered in drawn crystal units of good quality. It is obvious that this unit may be unstable at high temperatures, since the current transmission factor (α) is greater than unity at 70°C. The rise in α is the result of the very large increase in the hole density in the collector which accompanies the temperature rise.

In terms of the conventional tee network of r_e , r_b , r_c , and α , two points are interesting. First, the collector resistance $r_c \simeq 1/h_{22} \simeq 9$ megohms may seem very high. Actually, the lower values so commonly encountered are primarily the result of high leakage conductance, rather than a large electronic conductance. Second, the tee network base resistance $r_b \simeq h_{12}/h_{22} \simeq 1000$ ohms is nearly four times the high frequency feedback base resistance of 278 ohms.

5.3 *p-n-p Fused Junction Transistor*

The physical structure assumed for this unit is that of Fig. 2(b). The ring base contact is of particular importance. The material and structure are chosen to facilitate comparison with a specific development model for which a large amount of data is available. The emitter diameter is 15 mils and the collector diameter 30 mils. The base contact ring diameter is 40 mils.

The physical parameters of interest are:

V_c	Collector voltage, -4.5 volts
I_e	Emitter current, 1 ma
w_3	Wafer thickness, 3.5 mils
w_2	Collector to surface thickness, 2.1 mils
w_1	Collector to emitter thickness (not including depletion layers), 3.6×10^{-3} cm or 1.41 mil
ρ_b	Base resistivity, 1.5 ohm cm
τ_b	Hole lifetime in base region, 20 μ sec
ρ_e, ρ_c	Emitter and Collector resistivities unknown but very low, probably 0.001 ohm cm
τ_e, τ_c	Electron lifetime in emitter and collector unknown but very low, probably 0.1 μ sec
A_e	Emitter area, 0.00114 cm ²
A_c	Collector area, 0.00456 cm ²

The transmission parameters are then

$$g_{ee} = 0.039 \text{ mho}$$

$$\gamma_0 \text{ probably } \geq 0.995$$

$$\beta_0 = 0.993$$

$$\alpha_0^* \simeq 1.0000$$

$$a_{ee} \geq 0.988$$

$$\frac{\partial w}{\partial V_c} = \frac{x_m}{2V_c} = \frac{2.6 \times 10^{-4}}{2 \times 4.5} = 2.89 \times 10^{-5} \text{ cm/volt}$$

$$g_{cc} \simeq 8.15 \mu\text{mhos}$$

The parameters which determine frequency response are:

$$f_\alpha = \frac{\omega_\alpha}{2\pi} = \frac{D_p}{\pi w_1^2} = \frac{44}{\pi(3.6 \times 10^{-3})^2} = 1.08 \text{ mcps}$$

$$C_T = 5,000 \mu\mu\text{fd/cm}^2 \text{ for collector}$$

$$C_T \simeq 20,000 \mu\mu\text{fd/cm}^2 \text{ for emitter}$$

so that

$$C_{Tc} = 22.8 \mu\mu\text{fd}$$

$$C_{Te} = 22.8 \mu\mu\text{fd}$$

The effective hole storage capacitances of the emitter and collector are:

$$C_{se} = 3840 \mu\mu fd$$

$$C_{sc} = 0.8 \mu\mu fd$$

By the equations given in Section 3.0

$$r_{b'1} = 55 \text{ ohms}$$

$$r_{b'2} = 35 \text{ ohms}$$

The low frequency values of the “*h*” parameters are

$$h_{21} \simeq -0.988$$

$$h_{12} \simeq 2.09 \times 10^{-4}$$

$$h_{22} \simeq 0.17 \mu mho$$

$$h_{11} \simeq 26 + (0.012)55 \simeq 26.7$$

These theoretical values are compared with observed values in Table I. The major discrepancy in h_{21} is charged to surface recombination of injected holes, which was ignored in the calculation. It should be noted that h_{22} becomes $0.45 \times 10^{-6} \text{ mho}$ if the calculated value is corrected by the ratio $0.032/.012$, which is the ratio of the measured and calculated $(1 - \alpha)$'s or $(1 + h_{21})$'s. The difference between computed and measured h_{11} is the sum of a number of effects. First, the actual $(1 - \alpha)$ is greater than the computed value. Next, the junction temperature was probably greater than the assumed 25°C. Finally, the carrier injection level is high enough to modify the emitter diode properties in this direction.

The difference between calculated and observed current transmission cutoff is greater than appears from the data, since the theoretical three db response frequency is about 22 per cent higher than the “alpha

TABLE I

	Calculated	Measured
h_{21}	- 0.988	-0.968
h_{22}	$0.17 \times 10^{-6} \text{ mho}$	$0.48 \times 10^{-6} \text{ mho}$
h_{12}	2.09×10^{-4}	1.85×10^{-4}
h_{11}	26.7 ohm	33 ohm
C_e	22.8 $\mu\mu fd$	24.7 $\mu\mu fd$
f_{α}	1.08 mcps	0.95 mcps
$r_{b'1}$	55 ohm	55 ohm
$r_{b'2}$	35 ohm	63 ohm

cutoff" frequency calculated here. The difference is believed to be the result of the fact that holes emitted around the emitter periphery have much longer transit paths than do those emitted into the region directly between the electrodes and consequently reduce significantly the current cutoff frequency.

The serious discrepancy in $r_b'_{21}$ is probably the result of the $r_b'_{11}$ calculation being very pessimistic because of neglect of peripheral emission effects and of w_2 and w_3 being somewhat smaller than the assumed values. Again it can be seen that the equivalent tee base resistance $r_b = h_{12}/h_{22}$ is 375 ohms, nearly seven times the high frequency resistance of 55 ohms.

5.4 Qualitative Comparison

As might be expected, the qualitative agreement between theory and observation is better than the quantitative. For example, C_c , h_{22} , and h_{12} vary approximately as $(V_c')^{-1/2}$ in fused junction units. Alpha cutoff frequency increases as collector reverse bias is increased—a natural result of the narrowing of the base region. The qualitative discrepancies that are found are usually associated with large experimental deviations from the assumptions of the analysis.

5.5 Review of Design Calculations

Numerical analysis of both drawn junction and fused junction transistors has shown rather good agreement of theory and experiment. It is necessary, however, to modify some of the results empirically because of lack of full understanding of some effects, such as leakage and surface recombination.

Qualitative agreement of measurements and theory is, of course, much better than the quantitative correlation. For example, the dependence of all of the parameters on emitter current and collector voltage is almost exactly that expected from theory. Some of the dependence on emitter currents involves high carrier injection level theory which has been omitted from this study.

6.0 SUMMARY

6.1 Transmission Theory

Design theory of the small signal transmission parameters of junction transistors is relatively complete.

A one-dimensional analysis of minority carrier diffusion currents in terms of short circuit admittances has been combined with a similar analysis of depletion layer capacitances and an approximate three-

dimensional analysis of ohmic base region spreading resistance. The resulting equivalent circuit has characteristics in good agreement with experimental observations. In particular, collector capacitance, ohmic base region spreading resistance and the current transmission three db cutoff frequency may be computed with fair accuracy. The low frequency values of the common base hybrid parameters may also be calculated, but neglect of surface recombination and surface leakage results in serious errors in the short circuit current transmission factor h_{21} and the open circuit collector conductance h_{22} . The deviations of these two parameters from calculated values are, however, both reasonable and mutually consistent. The ohmic base layer spreading resistance, *which is the only base resistance of importance at high frequencies*, is very often much smaller than the low frequency base resistance appearing in an equivalent tee network.

Qualitative agreement of theory and measurement is excellent. The variation of all parameters with emitter current and collector voltage is within a few per cent that predicted from theory.

6.2 State of the Art

Since this paper was deliberately limited in scope, it is pertinent both to review its objectives and to point out significant omissions. The principal objective sought was presentation of small signal transmission design theory. No attempt was made to give a simple explanation of the junction transistor, relating both its large signal and its transmission characteristics to simple physical assumptions. While the design theory presented consolidates in one place some already published information, much remains to be done in assembling and integrating such knowledge from its present widely scattered locations.

In addition there exists a more detailed understanding of junction transistor characteristics than can be found in the literature. For example, units are found occasionally with negative h_{12} . This is a result of an easily modulated high resistance between base region and base contact (high μ_{bc}). Publication of such information can reduce by a few db the amount of head-scratching done by production engineers.

Other phenomena for which explanations have been developed are surface recombination and high carrier injection level effects. Despite this, much work remains to be done.

ACKNOWLEDGEMENTS

The general point of view taken here has been much influenced by discussions with J. A. Morton. Comments and criticisms by R. M. Ryder have been particularly helpful in the preparation of this material.

APPENDIX A

1.0 GENERAL

This study is an extension of Shockley's analysis of the junction transistor to include high-frequency effects and the voltage dependence of base-layer thickness. Shockley's paper* and the later paper by Shockley, Sparks, and Teal† contain the following of interest here:

- (a) analysis of the dc steady state of a junction transistor*;
- (b) analysis of the low-frequency small-signal parameters r_e , α , C_c^* †; and
- (c) analysis of frequency dependence of the transport factor β †.

In addition to repeating the above, this study gives these new results:

- (a) analysis of steady-state small-signal ac operation [dc biases present]; and
- (b) the small-signal ac short-circuit admittances y_{ee} , y_{ce} , y_{ec} , and y_{cc} .

1.1 ASSUMPTIONS

Semiconductor.

The p - n - p type structure assumed is shown in Fig. 6. The emitter, base, and collector regions may each be characterized by a resistivity and a minority carrier diffusion length. The emitter, collector, and base contacts have no effect on the currents which flow at the junctions. Injected carriers pass through the base layer by diffusion and through

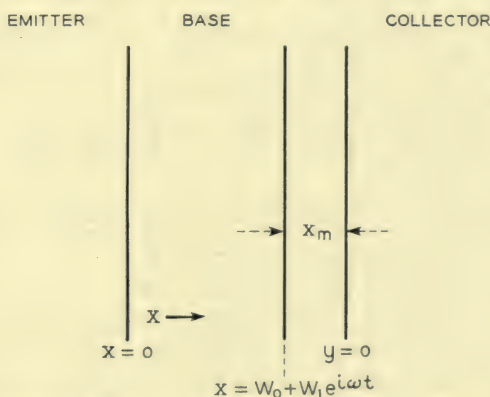


Fig. 6 — p - n - p junction transistor.

* W. Shockley, The Theory of p - n Junctions in Semiconductors and p - n Junction Transistors, B.S.T.J., **28**, p. 435.

† W. Shockley, M. Sparks, and G. K. Teal, The p - n Junction Transistors, Phys. Rev., **83**, p. 151, July, 1951.

the barrier layers* by drift. Base-layer thickness w depends on collector potential V_c through variation of the collector barrier thickness x_m . Variations of emitter barrier thickness are unimportant. The junctions are parallel.

Currents and Potentials

The currents and potentials studied are those at the collector and emitter barriers. Unless otherwise specified, "potential" implies difference in majority carrier Fermi levels, i.e., externally applied potential, rather than difference in electrostatic potential. The collector reverse potential is assumed to be many multiples of kT/q (e.g. > 0.5 volts), so that the classical p - n diode reverse conductance may be neglected. At 0.5 volt, this is of the order of 10^{-11} mhos. It decreases in magnitude one decade per sixty millivolts of bias potential.

Base Resistance

Majority carrier resistance in the base layer is not considered here.

Other Assumptions

Surface effects are excluded from consideration. In addition, several mathematical approximations of little physical consequence appear in the text as needed.

1.2 METHOD

The procedure employed is substantially that used by Shockley with some additions. First, minority carrier concentrations on both sides of each barrier are related to the barrier potentials.

The dc minority carrier distribution in each of the three transistor regions is then computed from these boundary conditions with the aid of the continuity equation.

Next, small-signal ac perturbations of the barrier potentials and of the minority carrier densities at the barriers are used in the same way to find ac distributions of the minority carriers. The effects of voltage dependence of base-layer thickness are found by means of a small-signal ac perturbation of the position of the collector side of the base layer. The resulting ac distribution of minority carriers is computed as before.

Finally, dc and ac currents at emitter and collector barriers are com-

* I.e., charge depletion layers.

puted from the gradients of the minority carrier densities at the barriers.

In the ac analysis, the forward rotating time function $e^{i\omega t}$ is used. In general, the first-order solution for small signals is obtained by assuming that the disturbance associated with $e^{i\omega t}$ is small, by neglecting its powers and harmonics, and by using first-order expansions, e.g.,

$$e^{\frac{w_0}{L} + \frac{w_1}{L} e^{i\omega t}} \approx \left(1 + \frac{w_1}{L} e^{i\omega t}\right) e^{\frac{w_0}{L}}$$

The ac magnitudes such as w_1 , p_{e1} , n_{c1} , represent complex phasors of the form $ae^{j\phi}$. A list of symbols is given in Section 1.5.

1.3 ANALYSIS

In the base layer at $x = 0$,

$$p = p_{e0} + p_{e1}e^{i\omega t} = p_n e^{qV_e/kT} \quad (1a)$$

at $x = w_0$,

$$p = p_{e0} + p_{c1}e^{i\omega t} = p_n e^{qV_c/kT} \quad (1b)$$

in which

$$V_e = V_{e0} + V_{e1}e^{i\omega t}$$

$$V_c = V_{c0} + V_{c1}e^{i\omega t}$$

and $V_{e1} \ll V_{e0}$ and $V_{c1} \ll V_{c0}$ so that

$$p_{e0} \approx p_n e^{qV_{e0}/kT} p_{c1} \approx p_{e0} \frac{q}{kT} V_{c1}$$

$$p_{c0} \approx p_n e^{qV_{c0}/kT} p_{c1} \approx p_{c0} \frac{q}{kT} V_{c1}$$

P_{e0} , P_{c0} , V_{e0} , and V_{c0} are average values while p_{e1} , p_{c1} , V_{e1} , and V_{c1} are ac phasors.

The continuity equation for holes for one-dimensional flow is:

$$D_p \frac{\partial^2 p}{\partial x^2} - \left(\frac{p_n - p}{\tau} \right) = \frac{\partial p}{\partial t} \quad (2)$$

in which τ is hole lifetime. A solution to equation (8-18) is:

$$p = p_n + Ae^{x/L} + Be^{-x/L} + Ce^{sx/L + i\omega t} + De^{-sx/L + i\omega t} \quad (3)$$

where $L = \sqrt{D\tau}$ and $s = (1 + i\omega\tau)^{1/2}$.

Application of the boundary values of equation (1a) and (1b) to

equation (3) gives hole density in the base layer:

$$\begin{aligned}
 p_0(t, x) = p_n + & \left[\frac{(p_{c0} - p_n) - (p_{c0} - p_n)e^{-w_0/L}}{2 \sinh(w_0/L)} \right] e^{x/L} \\
 & - \left[\frac{p_{c0} - p_n - (p_{c0} - p_n)e^{w_0/L}}{2 \sinh(w_0/L)} \right] e^{-x/L} \\
 & + \left[\frac{p_{c1} - p_{e1}e^{-sw_0/L}}{2 \sinh(sw_0/L)} \right] e^{sx/L + i\omega t} \\
 & - \left[\frac{p_{c1} - p_{e1}e^{sw_0/L}}{2 \sinh(sw_0/L)} \right] e^{-sx/L + i\omega t}
 \end{aligned} \tag{4}$$

Since up to this point w has been assumed constant [$w = w_0$], equation (4) does not include effects of voltage dependence of base layer thickness. To introduce these, a new set of boundary conditions is used: at $x = 0$,

$$\text{and at } x = w_0 + w_1 e^{i\omega t}, \quad p = p_{c0} + p_{e1}e^{i\omega t} \tag{5a}$$

$$p = p_{c0} + p_{c1}e^{i\omega t} \tag{5b}$$

in which $w_1 \leq w_0$ and is a phasor.

$$w_1 = \frac{\partial w}{\partial V_c} V_{c1}$$

It can be seen that conditions are as before except that the collector side of the base layer swings about position w_0 at angular frequency ω .

A solution of equation (31) with conditions (5a) and (5b) is given by

$$p_1(t, x) = p_0(t, x) + \tilde{p}(t, x) \tag{6}$$

in which $p_0(t, x)$ is given by equation (4) and $\tilde{p}(t, x)$ is the perturbation associated with $w_1 e^{i\omega t}$.

If equations (5a) and (5b) are rewritten in terms of $\tilde{p}(t, x)$, they become, using first order expansions:

$$\tilde{p}(t, 0) = 0 \tag{7a}$$

$$\begin{aligned}
 \tilde{p}(t, w_0 + w_1 e^{i\omega t}) = & [(p_{c0} - p_n) \operatorname{csch}(w_0/L) \\
 & - (p_{c0} - p_n) \coth(w_0/L)] \frac{w_1}{L} e^{i\omega t}
 \end{aligned} \tag{7b}$$

Since $\tilde{p}(t, x)$ is an ac solution of the continuity equation (2), it has the form:

$$\tilde{p}(t, x) = E e^{sx/L + i\omega t} + F e^{-sx/L + i\omega t} \tag{8}$$

Use of equations (7a) and (7b) leads to:

$$p(t, x) = \frac{\sinh (sx/L)}{\sinh (sw_0/L)} [(p_{e0} - p_n) \operatorname{csch} (w_0/L) - (p_{e0} - p_n) \coth (w_0/L)] \frac{w_1}{L} e^{i\omega t} \quad (9)$$

The complete solution for hole density in the base layer is:

$$\begin{aligned} p_1(t, x) = p_n + & \left[\frac{(p_{e0} - p_n) - (p_{e0} - p_n)e^{-w_0/L}}{2 \sinh (w_0/L)} \right] e^{x/L} \\ & - \left[\frac{(p_{e0} - p_n) - (p_{e0} - p_n)e^{w_0/L}}{2 \sinh (w_0/L)} \right] e^{-x/L} \\ & + \left[\frac{p_{e1} - p_{e1}e^{-sw_0/L}}{2 \sinh (sw_0/L)} \right] e^{sx/L + i\omega t} \\ & - \left[\frac{p_{e1} - p_{e1}e^{sw_0/L}}{2 \sinh (sw_0/L)} \right] e^{-sx/L + i\omega t} \\ & + \frac{w_1}{L} e^{i\omega t} \frac{\sinh (sx/L)}{\sinh (sw_0/L)} [(p_{e0} - p_n) \operatorname{csch} (w_0/L) \\ & - (p_{e0} - p_n) \coth (w_0/L)] \end{aligned} \quad (10)$$

The hole-current density in the base layer is found from equation (10) by the use of the equation for diffusion current

$$I_p = -qD_p \frac{dp}{dx} \quad (11)$$

which yields

$$\begin{aligned} I_p = -q \frac{Dp}{L} & \left((p_{e0} - p_n) \frac{\cosh (x/L)}{\sinh (w_0/L)} - (p_{e0} - p_n) \cdot \right. \\ & \left. \frac{\cosh [(x - w_0)/L]}{\sinh (w_0/L)} + sp_{e1} \frac{\cosh (sx/L)}{\sinh (sw_0/L)} e^{i\omega t} \right. \\ & \left. - sp_{e1} \frac{\cosh [(sx - sw_0)/L]}{\sinh (sw_0/L)} e^{i\omega t} + s \frac{w_1}{L} e^{i\omega t} \frac{\cosh (sx/L)}{\sinh (sw_0/L)} \right. \\ & \left. [(p_{e0} - p_n) \operatorname{csch} (w_0/L) - (p_{e0} - p_n) \coth (w_0/L)] \right) \end{aligned}$$

Hole-current densities at emitter and collector may be found by substitution of $x = 0$ and $x = w_0$ respectively. The first two terms in equation (12) give dc current components* which may be attributed to

* With some labor, these terms may be converted to Shockley's equation (5.6).

collector and emitter potentials in that order, while the last three are ac terms resulting from collector potential, emitter potential, and base layer thickness variation respectively.

The hole-current densities may be related to the ac potentials by use by use of the approximations given at the beginning of paragraph 1.3.

$$p_{e1} \approx p_{e0} \frac{q}{kT} V_{e1} \qquad p_{c1} \approx p_{c0} \frac{q}{kT} V_{c1}$$

and the relation

$$w_1 = \frac{\partial w}{\partial V_c} V_{c1}$$

However, the average collector potential V_{c0} is negative and many times kT/q so that $p_{c0} \simeq 0$ and $p_{c1} = 0$. The ac emitter hole-current density is therefore

$$\begin{aligned} I_{pe1} e^{i\omega t} = \frac{qDp}{L} \left(V_{e1} \frac{q}{kT} s p_{e0} \coth (sw_0/L) \right. \\ \left. - \frac{V_{e1}}{L} \frac{\partial w}{\partial V_c} s \operatorname{csch} (sw_0/L) [(p_{e1} - p_n) \operatorname{csch} (w_0/L) \right. \\ \left. - (q_{c0} - p_n) \coth (w_0/L)] \right) e^{i\omega t} \end{aligned} \quad (13a)$$

If $p_{e0} > p_n$, equations (13a) and (13b) becomes very closely

$$I_{pe1} = \left[I_{pe0} \frac{q}{kT} \frac{s \tanh (w_0/L)}{\tanh (sw_0/L)} V_{e1} + I_{pc0} \frac{\partial w}{\partial V_c} \frac{s V_{c1}}{L \sinh (sw_0/L)} \right] \quad (13b)$$

In equation (13b), I_{pe0} and I_{pc0} are average emitter and collector hole-current densities. The entire coefficients of V_{e1} and V_{c1} in equation (13b) are the input and the feedback short-circuit admittances associated with hole flow in the transistor.

Similarly, forward transfer and output short-circuit admittances associated with hole flow may be found from equation (12) by substitution of $x = w_0$ and use of the approximation $p_1 \simeq p_1 - p_n$, i.e., $p_1 \gg p_n$. In calculating collector current, the sign of equation (12) must be reversed, since equation (12) gives current flow in the x -direction while collector current is assumed to flow in the negative x -direction. The admittances are given in the summary at the end.

Next, there are admittances associated with electron flow in the p - n - p transistor. Flow of electrons from base to emitter gives rise to an input admittance term, while electrons flowing from collector to base give rise to output and forward transfer admittance terms. An outline

of the derivation of these terms follows. The terms are given in the summary.

For electrons in the emitter:

at $x = 0$,

$$n_e = n_{e0} + n_{e1}e^{i\omega t} = n_p e^{qV_e/kT}$$

at $x = -\infty$,

$$n_e = n_p$$

in which, again, $V_e = V_{e0} + V_{e1}e^{i\omega t}$ and first-order expansions are used. $x = -\infty$, $n_e = n_p$, is chosen as a boundary condition in order to eliminate effects of the emitter contact. Solution of the continuity equation results in:

$$n_{0e}(t, x) = n_p + (n_{0e} - n_p) e^{x/L} + n_{01} e^{-sx/L + i\omega t} \quad (14)$$

in which the τ is $s = (1 + i\omega\tau)^{1/2}$ now implies electron lifetime in the emitter body. Electron diffusion current at the emitter is computed by means of the equation corresponding to equation (11), giving

$$I_{ne} = \frac{qD_n}{L_n} ((n_{e0} - n_p) + n_{e1}e^{i\omega t}) \quad (15)$$

The ac admittance associated with the last term appears in the summary, Section 1.4.

The ac electron current from the collector is not a diffusion current, but rather a drift current resulting from the hole current flowing in the collector body. Since the ac electron current is directly proportional to the ac hole current in the collector, the result is an effective multiplication of the output and forward transfer admittances associated with hole current in the collector body.*

For electrons in the collector the boundary conditions are:

at $y = 0$,

$$n = n_p e^{qV_c/kT} = 0 \quad (16a)$$

at $y = \infty$,

$$n = n_p \quad (16b)$$

The condition $n = 0$ at the edge of the barrier region results from V_c .

* Space-charge layer widening effects are neglected since they are usually very small and are difficult to analyze.

being negative and many times kT/q . The distribution of electrons in the collector is governed by conditions (16a) and (16b) and the modified continuity equation* developed by W. van Roosbroeck.

$$\frac{\partial n}{\partial t} = -\frac{n - n_p}{\tau} + \frac{\mu_n E_a}{M_0} \frac{\partial n}{\partial y} + D_0 \frac{\partial^2 n}{\partial y^2} \quad (17)$$

in which E_a is the electric field which would be associated with the total current if $n \equiv n_p$. $M_0 = (p_p + bn_0)/(p_p + n_p)$ and $D_0 = (p_p + n_p)/(p_p/D_n + n_p/D_p)$.

If n is assumed to be given by $n(t, x) = n_0(y) + n_1(y)e^{i\omega t}$ and E is assumed to be

$$E_a(t, x) = E_0 + E_1 e^{i\omega t} = \rho_c I_{c0} + \rho_{pc} I_{pc1} e^{i\omega t}$$

in which I_{c0} is average collector current and I_{pc1} is the ac hole current of the collector, equation (17) may be reduced to two equations

$$0 = -\left(\frac{n_0 - n_p}{D_0 \tau}\right) + \frac{\mu_n E_0}{D_0 M_0} \frac{\partial n_0}{\partial y} + \frac{\partial^2 n_0}{\partial y^2} \quad (18a)$$

$$-\frac{\mu_n E_1}{M_0 D_0} \frac{\partial n_0}{\partial y} = -\frac{1 + i\omega \tau}{D_0 \tau} n_1 + \frac{\mu_n E_0}{D_0 M_0} \frac{\partial n_1}{\partial y} + \frac{\partial^2 n_1}{\partial y^2} \quad (18b)$$

The solution of equation (18a) under conditions (16a) and (16b) is

$$n_0 = n_p(1 - e^{m_2 y}) \quad (19)$$

where

$$m_2 = -\frac{\mu_n E_0}{2D_0 M_0} - \sqrt{\left(\frac{\mu_n E_0}{2D_0 M_0}\right)^2 + \frac{1}{D_0 \tau}}$$

Inspection of equation (19) shows that

$$\frac{\partial n_0}{\partial y} = -m_2 n_p e^{m_2 y} \quad (20)$$

It is then apparent that n_1 must be of the form

$$n_1 = n_p e^{m_2 y} f(y) \quad (21)$$

Substitution of equation (2) and equation (21) into (18b) leads to

$$\frac{\partial^2 f}{\partial y^2} + \left(m_2 + \frac{\mu_n E_0}{M_0 D_0}\right) \frac{\partial f}{\partial y} - \frac{i\omega}{D_0} f = \frac{n_p m_2 \mu_n e_2}{M_0 D_0} \quad (22)$$

A general solution of equation (22) is

$$f(y) = \frac{n_p m_2 \mu_n E_1}{-i\omega M_0} + B e^{\tau_1 y} + C e^{\tau_2 y} \quad (23)$$

* W. van Roosbroeck; private communication.

where

$$r_1, r_2 = -\frac{1}{2} \left(m_2 + \frac{\mu_n E_0}{D_0 M_0} \right) \pm \sqrt{\left[\frac{1}{2} \left(m_2 + \frac{\mu_n E_0}{D_0 M_0} \right) \right]^2 + \frac{i\omega}{D_0}}$$

It may be shown that the boundary conditions on $f(y)$ are:

$$\text{at } y = 0, \quad f = 0; \quad \text{at } y = \infty, \quad f = 0.$$

Application of these values to equation (23) results in

$$f(y) = \frac{n_p m_2 \mu_n E_1}{i\omega M_0} [1 - e^{r_2 y}] \quad (24)$$

in which r_2 is obtained using the negative square root.

The electron density in the collector is now given by

$$n = n_p (1 - e^{m_2 y}) + \frac{n_p m_2 \mu_n E_1}{i\omega M_0} [1 - e^{r_2 y}] e^{m_2 y} e^{i\omega t} \quad (25)$$

The electron current is given by

$$I_n = n_p q D_n \left[-m_2 e^{m_2 y} + e^{i\omega t} \frac{\mu_n E_1 m_2}{i\omega M_0} (m_2 - m_2 e^{r_2 y} - r_2 e^{r_2 y}) e^{m_2 y} \right]. \quad (26)$$

At the collector junction, the ac component reduces to

$$I_{nc1} = \frac{n_p q D_n \mu_n m_2 r_2 E_1}{i\omega M_0} \quad (27)$$

Now, since $E_1 = \rho_{pc} I_{pc1}$, the collector multiplication factor

$$(I_{pc1} + I_{nc1})/I_{pc1}$$

is

$$\alpha^* = 1 + \frac{n_p q D_n \mu_n m_2 r_2}{i\omega M_0} \rho_{pc} \quad (28a)$$

$$\alpha^* = 1 + \frac{\sigma_{nc}}{\sigma_{pc}} \frac{D_n m_2 r_2}{i\omega M_0} \quad (28b)$$

The effect of collector multiplication as given in equations (28a) and (28b) is included in the general admittance expressions given later.

Finally, no mention has been made of the admittances associated with barrier capacitances. Since the currents which charge these are majority carrier currents, there are no input-output interactions except those associated with majority carrier resistance of the base layer [an effect not analyzed in this study]. These capacitances add directly to input and output admittances. Shockley* gives methods for calculating these capacitances.

* Loc. cit., vol. 28, page 435.

1.4 SUMMARY

In paragraph 1.3, derivations were given or outlined for each of the four small-signal short-circuit admittances associated with the hole flow, electron flow, and barrier capacitances in junction transistors. The terms appear in that order in the expressions which follow.

$$y_{ee} = \left[\frac{qp_n D_p s_p}{L_p} \coth (s_p w_0 / L_p) + qn_p D_n s_{ne} / L_{ne} \right] \frac{q}{kT} e^{qV_{e0}/kT} + i\omega C_{Te} \quad (29)$$

$$y_{ce} = - \left[q \frac{p_n D_p s_p}{L_p} \operatorname{csch} (s_p w_0 / L_0) \right] \frac{q}{kT} e^{qV_{e0}/kT} \left(1 + \frac{\sigma_{nc}}{\sigma_{pc}} \frac{D_n m_2 r_2}{i\omega M_0} \right) \quad (30)$$

$$y_{ec} = - \frac{\partial w}{\partial V_c} \frac{s_p}{L_p \sinh (s_p w_0 / L_p)} \frac{qp_n D_p}{L_p} [(e^{qV_{e0}/kT} - 1) \operatorname{csch} (w_0 / L_p) + \coth (w_0 / L_p)] \quad (31)$$

$$y_{cc} = \left(\frac{\partial w}{\partial V_c} \frac{s_p}{L_p \tanh (s_p w_0 / L_p)} \frac{qp_n D_p}{L_p} [(e^{qV_{e0}/kT} - 1) \operatorname{csch} (w_0 / L_p) + \coth (w_0 / L_p)] + i\omega C_{Te} \right) \left(1 + \frac{\sigma_{nc} D_n m_2 r_2}{\sigma_{pc} i\omega M_0} \right) \quad (32)$$

in which all symbols are defined in Section 1.5. It should be noted that collector multiplication operates on the current to the barrier capacitance since the latter current is a hole current in the collector body.

The term $\partial w / \partial V_c$ is the same for both p - n - p and n - p - n structures. It is: for step junctions

$$\frac{\partial w}{\partial V_c} = - \frac{x_m}{2V_c} \quad (33)$$

and for graded junctions

$$\frac{\partial w}{\partial V_c} = - \frac{x_m}{6V_c} \quad (34)$$

V_c in equations (33) and (34) means dc electrostatic potential difference across the collector barrier.

Equations (29) through (32) may be manipulated into many forms. One of these sets which may be employed as a starting point for the approximate forms given in the body of this chapter is:

$$y_{ee} = \frac{q}{kT} \left[I_{pe0} \frac{s_p \tanh (w_0 / L_p)}{\tanh (s_p w_0 / L_p)} + I_{ne s_{se}} \right] + i\omega C_{Te} \quad (35)$$

$$y_{ce} = -\frac{q}{kT} I_{pe0} \frac{s_p \tanh(w_0/L_p)}{\sinh(s_p w_0/L_p)} \left[1 + \frac{\sigma_{nc}}{\sigma_{pc}} \right]^* \quad (36)$$

$$x_{cc} = \frac{\partial w}{\partial V_c} \frac{s_p}{L_p \sinh(s_p w_0/L_p)} I_{pe0} \quad (37)$$

$$y_{cc} = \left[\frac{-\partial w}{\partial V_c} \frac{s_p}{L_p \tanh(s_p w_0/L_p)} I_{pe0} + i\omega C_{Te} \right] \left[1 + \frac{\sigma_{nc}}{\sigma_{pc}} \right]^* \quad (38)$$

The change in signs which occurs in going from equations (31) and (32) to equations (37) and (38) takes place because the current replaced by I_{pc} had the opposite assigned positive sense.

1.5 SYMBOLS USED IN THE APPENDIX

C_c, C_{Tc} = collector barrier capacitance.

C_{Te} = emitter barrier capacitance.

D_n, D_p = diffusion constants for electrons and holes

$D_0 = (p_p + n_p)/(p_p/D_n + n_p/D_p)$

$E_a = ip$ = electric field associated with current at thermal equilibrium carrier densities.

E_1, E_0 = ac and dc components of E_a

$I_{ne}, I_{ne0}, I_{ne1}, I_{nc}, I_{nc0}, I_{nc1}$ = total, average, and ac emitter and collector electron currents

$I_{pe}, I_{pe0}, I_{pe1}, I_{pc}, I_{pc0}, I_{pc1}$ = total, average, and ac emitter and collector hole currents

kT/q = thermal energy of carriers = 0.026 electron-volt

L_p, L_n = diffusion lengths for holes and electrons

$m_2 = -\frac{\mu_n E_0}{2D_0 M_0} - \sqrt{\left(\frac{\mu_n E_0}{2D_0 M_0}\right)^2 + \frac{1}{D_0 \tau}}$ decay constant for average electron density in the collector

$M_0 = (p_p - bn_p)/(p_p = n_p)$

n_p, n_n = thermal equilibrium electron densities in p and n regions

n_{e0}, n_{e1} = dc and ac components of electron density at emitter junction

p_n, p_p = thermal equilibrium hole densities in n and p regions

$p_{e0}, p_{e1}, p_{c0}, p_{c1}$ = dc and ac components of hole density at emitter and collector junctions

q = electronic charge, 1.6×10^{-19} coulombs

r_e = ac emitter resistance

$$r_1, r_2 = -\frac{1}{2} \left(m_2 + \frac{\mu_n E_0}{D_0 M_0} \right) \pm \sqrt{\left[\frac{1}{2} \left(m_2 + \frac{\mu_n E_0}{D_0 M_0} \right) \right]^2 + \frac{i\omega}{D_0}}$$

* The factor $a^* = (1 + \sigma_{c/\sigma_{pc}})$ is current dependent. At small average collector currents, it is $(1 + \sigma_{nc}/2\sigma_{pc})$ and rises to $(1 + \sigma_{nc}/\sigma_{pc})$ at high current densities.

$$s = (1 + i\omega\tau)$$

T = temperature in degrees Kelvin (absolute temperature)

$V_e, V_{e0}, V_{e1}, V_c, V_{c0}, V_{c1}$ = total, average, and ac emitter and collector potentials.

V_c = average collector potential

w, w_0, w_1 = total, average, and ac variation of base-layer thickness

x = distance from emitter barrier

x_m = thickness of collector barrier layer

y = distance from collector side of collector barrier layer

$y_{ee}, y_{ce}, y_{ec}, y_{cc}$ = short-circuit ac admittances

α = current amplification factor

β = base-layer transport factor

ρ_c = resistivity of collector region

ρ_{pc} = resistivity for holes in collector region

σ_{pc}, σ_{nc} = conductivities for holes and electrons in collector region

τ = minority carrier lifetime

μ_n = electron mobility

μ_p = hole mobility

ω = angular frequency in radians

APPENDIX B

BASE LAYER SPREADING RESISTANCE

Significance

The bulk resistance of the base layer or base layer spreading resistance (r_b') is important because base current passing through r_b' produces a base contact to emitter junction negative feedback voltage $i_b r_b'$. Reduction of r_b' is an important objective in improvement of junction units.

Types

Since feedback voltage to the emitter junction is produced by two separately measurable base current components having very different flow paths, two separately measurable base layer spreading resistances ($r_{b'1}$ and $r_{b'2}$) may be defined.

The current $(1-\alpha)i_e$ originates in the base layer between the emitter and collector junctions and flows radially through the base layer to the base contact producing a feedback voltage at the emitter; $r_{b'1}$ is defined as the ratio of this feedback voltage to the current.

The collector capacitance current $j\omega C_{ec}$ enters the base layer uni-

formly over the entire area of the collector junction resulting in a feedback voltage at the emitter; $r_b'2$ is defined as the ratio of this feedback voltage to this current.

Calculations

The resistances $r_b'1$ and $r_b'2$ for the transistor of Fig. 2(b) may be computed with the help of three formulas which give the feedback voltages for the three geometrical problems involved in this transistor. Each expression gives the voltage V developed at electrode C by a current I entering through electrode A and leaving through electrode B. The formulas are in terms of sheet resistance ρ_b/w (resistance per square in ohms) and the radii involved.

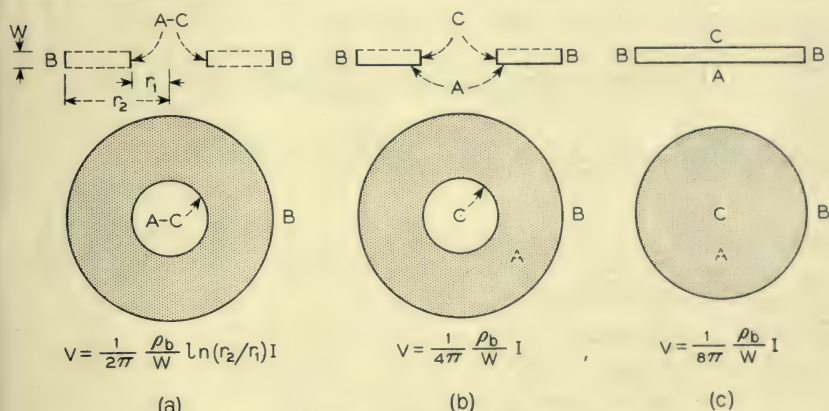


Fig. 7 — Feedback voltage for three geometries.

The simplest situation and its formula are shown in Fig. 7(a). Electrodes A and C are the same and expression gives the resistance of an annular ring to radial current flow.

Fig. 7(b) also shows an annular ring, but the current I is introduced uniformly over one of the flat surfaces, while the voltage V is measured from outside edge to inside edge.

Fig. 7(c) shows current introduced uniformly over the surface of a disc, while voltage V is the average voltage developed along the surface of the disc. It should be clear that electrodes A and C do not conduct parallel to the disc surface (i.e., they are not equipotentials).

Example

The r_b' 's for the transistor of Fig. 2(b) will be calculated.

A. $r_b'1$ — The current $I = (1-\alpha)i_e$ is assumed to originate uniformly

in the region between emitter and collector in Fig. 2(b). It flows radially outward to the base contact.

$$V = I \left[\frac{1}{8\pi} (\rho_b/w_1) + \frac{1}{2\pi} (\rho_b/w_2) \ln (r_2/r_1) + \frac{1}{2\pi} (\rho_b/w_3) \ln (r_3/r_2) \right]$$

$$r'_{b1} = V/I = \rho_b \left[\frac{1}{8\pi w_1} + \frac{1}{2\pi w_2} \ln \frac{r_2}{r_1} + \ln \left(\frac{r_3}{r_2} \right) \right]$$

B. r'_{b2} —Inspection of Fig. 7 shows that the collector capacitance current originating opposite the emitter junction (I_1) has the same flow path as the $(1-\alpha)i_e$ current. The remainder of the collector current which enters the base from the collector outside the emitter (I_2) likewise flows radially out to the base contact. The total feedback voltage developed between emitter junction and base contact is

$$V = I_1 r'_{b1} + I_2 \left[\frac{1}{4\pi} (\rho_b/w_2) + \frac{1}{2\pi} (\rho_b/w_3) \ln (r_3/r_2) \right]$$

$$r'_{b2} = \frac{V}{I_1 + I_2} = \rho \left\{ \frac{1}{2\pi w_3} \ln \left\{ \frac{r_3}{r_2} \right\} + \left\{ \frac{r_1}{r_2} \right\}^2 \left[\frac{1}{2\pi w_2} \ln \left\{ \frac{r_2}{r_1} \right\} + \frac{1}{8\pi w_1} \right] + \left[1 - \left\{ \frac{r_1}{r_2} \right\}^2 \right] \frac{1}{4\pi w_2} \right\}$$

C. Numerical Results

For $\rho_b = 1.5$ ohm-cm

$$w_1 = 0.5 \text{ mil}$$

$$w_2 = 1.0 \text{ mil}$$

$$w_3 = 2.0 \text{ mil}$$

$$r_2/r_1 = 2$$

$$r_3/r_2 = 2$$

$$r'_{b1} = 147 \text{ ohms}$$

$$r'_{b2} = 96 \text{ ohms}$$

Transistor Oscillator for Use in Multifrequency Pulsing Current Supply

By F. E. BLOUNT

(Manuscript received June 30, 1953)

This paper covers the design and performance of an oscillator using a transistor, in the multifrequency pulsing of digital information over telephone transmission media. The frequencies used are in the range from 700 to 1700 cycles per second.

INTRODUCTION

A large proportion of the telephone calls made require connections to be set up in more than one central office. Where common control systems are used, this requires that information needed by the second central office be transmitted to it from the first central office which in turn has received its information from the calling subscriber. The "language" used in some cases for transmitting the information is in the form of short pulses of alternating current. Each pulse consists of a combination of two of six available frequencies. Twelve combinations of the six frequencies make up the total "vocabulary". Ten are required for digits and two for special signals sent at the beginning and end of pulsing. This is known as multifrequency, MF, pulsing. The device which controls the pulses is called an MF sender. The device which receives the MF pulses and translates the information received for use by other equipment is called an MF receiver.

An operator may also use multifrequency pulsing when transmitting information by means of a key set to a distant office. Control of the frequencies used is obtained by contacts on the key depressed. Twelve keys are used.

The previous source of ac was a circuit capable of supplying both operator positions and senders. This equipment is however quite expensive. If only a few senders in an office require the MF current the cost per

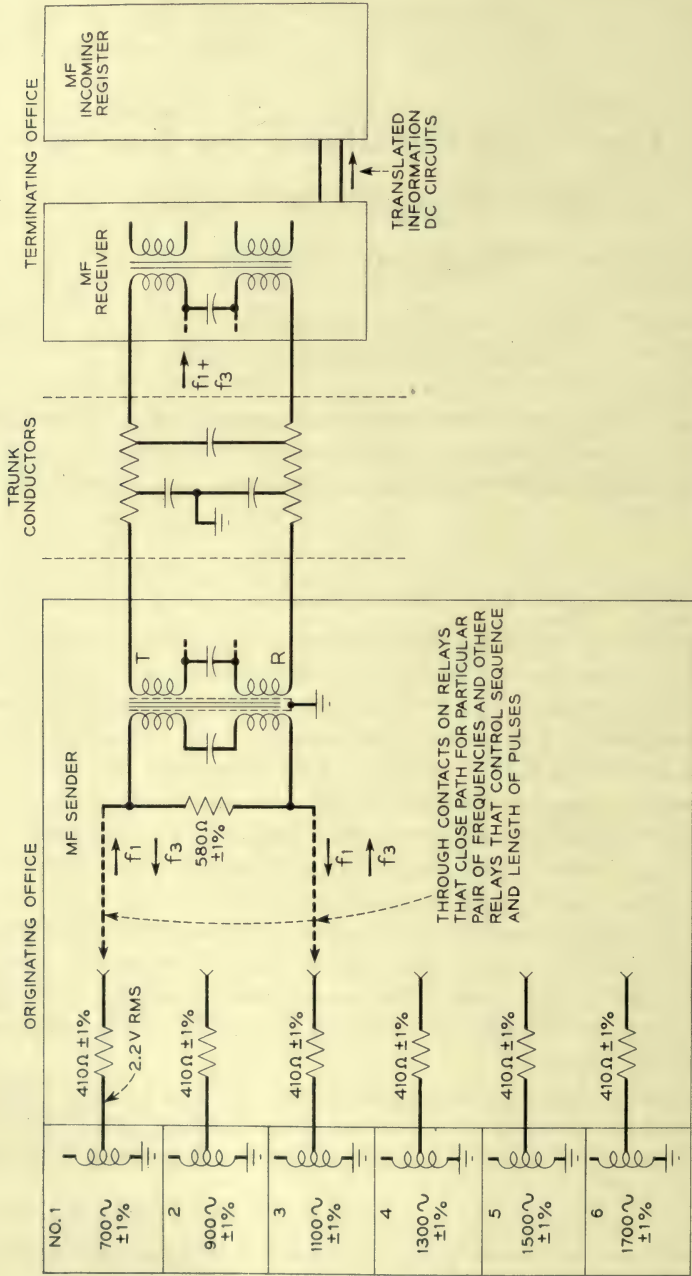


Fig. 1 — Simplified diagram of MF pulsing circuits.

sender becomes very large. It was for this reason that transistor oscillators were developed for use as individual current supplies.

An individual sender will handle a large number of calls during a day but the aggregate time during which the multifrequency current is required is normally less than two hours. The ability of transistors to operate the instant that power is applied is therefore a distinct advantage in saving power. The low voltage that is required for their operation and the long life expectancy are further advantages.

The method used is illustrated in Fig. 1. To transmit a pulse the output windings of two oscillators are connected in series through two 410-ohm resistors and the parallel combination of a 580-ohm resistor and windings on a transformer. The other windings on the transformer are connected to the trunk conductors. At the distant office the trunk is terminated by a similar transformer. The output of this transformer is connected to an amplifier, filter circuits and detectors in an MF receiver. Relays associated with the detectors convert the ac signal into a dc signal for the register circuit.

The 580-ohm resistor across the transformer winding serves as a termination for the trunk between pulses. This resistor in combination with the two 410-ohm resistors in the oscillator circuits serve as the trunk termination during pulsing.

REQUIREMENTS FOR OSCILLATOR

This oscillator was designed to be used in place of an existing piece of equipment. It therefore must have the properties of the existing equipment on which the design of equipments that are associated with it have been predicated.

The requirements are then:

1. Operate at 700, 900, 1100, 1300, 1500 or 1700 cycles per second $\pm$ 1 per cent.
2. Furnish ac signals at a level of $-3 \text{ dbm} \pm 1 \text{ db}$ at the input to the trunk.
3. The level difference at all frequencies must be no greater than 1 db.
4. Attain normal output level between the time the sender is seized and the time the first pulse is sent. This time is approximately one half second.
5. Meet all requirements with dc input voltage limits of 45–50 volts.
6. Meet all requirements with ambient temperatures of $+40^\circ\text{F}$ to $+135^\circ\text{F}$.
7. Have level of second harmonic a minimum of 35 db below the fundamental.

8. Provide an output impedance sufficiently low that the differences in impedance of the associated oscillator will not appreciably affect either output level or frequency.

9. Meet frequency requirements when working into trunks that have impedances with phase angles of $+15^\circ$ to -45° .

10. Be unaffected by transient disturbances on trunk conductors coupled to the oscillator. These disturbances are caused by lightning or troubles on closely associated power lines.

DISCUSSION

There are many ways of making a transistor oscillator. One of the simplest and most stable of these uses an LC circuit for controlling the frequency. Additional windings inductively coupled to the tuned circuit inductance provides the necessary feedback to the transistor and the connection to the load. This is the type of circuit adopted for this application. A schematic of the circuit is shown in Fig. 2. The transistor controls power supplied to the oscillatory circuit during only one half the operating cycle. The amplitude of the ac voltage produced across winding 1-2 is almost equal to the dc voltage applied to the collector.

The transistor acts as an amplifier or it may be considered as a negative resistance. Control of the transistor is obtained by current flowing in the emitter that is in phase with the collector voltage. To build up oscillations, the power put into the circuit must be greater than that expended. Stability is obtained when the power put in equals the power expended.

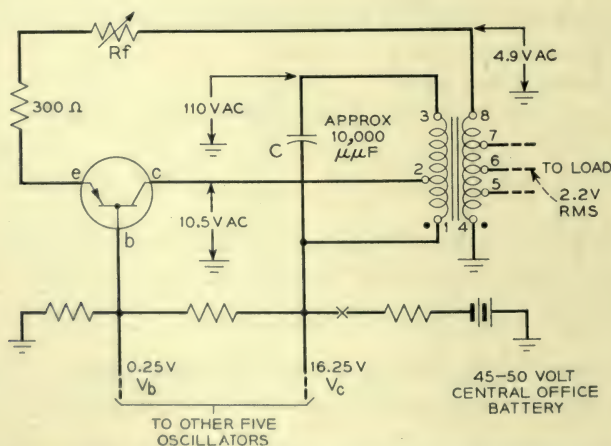


Fig. 2 — Oscillator circuit for MF signaling showing approximate values for voltages.

In the following sections the functions of a resonant circuit are first discussed. Then, the characteristics of a point contact transistor that make it suitable for use in an oscillator are considered. This is followed by a discussion of the behavior of the two when combined to form an oscillator with consideration given to the effect various factors have on performance.

RESONANT CIRCUIT

A resonant circuit has several properties of interest when one considers it as part of an oscillator. These are:

1. The input impedance of a parallel circuit consisting of an inductance, L , and a capacitance, C , has the property of a resistance at approximately frequency,¹

$$f_0 = \omega_0/2\pi.$$

That is, input impedance,

$$Z = RQ^2 \left(1 - j \frac{1}{Q} \right),$$

where R = effective resistance of the inductor and

$$Q = \frac{\omega_0 L}{R}$$

when Q is high, $Z \doteq RQ^2 = R_0$ the equivalent resistance at resonance.

2. The decrement of a parallel resonant circuit when disconnected from sources of power or loads is:²

$$\left[\begin{array}{c} \text{numerical decrement in cur-} \\ \text{rent or voltage per cycle} \end{array} \right] = 1 - e^{-\frac{R}{2Lf_0}}. \quad (1)$$

When the effective resistance R becomes zero the decrement likewise becomes zero.

In the oscillatory circuit considered the decrement is made zero by the use of a negative resistance equal to the effective positive resistance of the tuned circuit. Additional loads on the tuned circuit are accommodated in the same way.

If the negative resistance in the simple circuit considered becomes greater than the equivalent resistance of the tuned circuit and its loads, power will be supplied by the tuned circuit at a rate that will maintain equality between the power put into the circuit and the power taken out. As a corollary to this, if the negative resistance is less than the positive resistance, the voltage will increase across the circuit. For

the oscillator this would be:

$$\left[\begin{array}{c} \text{numerical increment in cur-} \\ \text{rent or voltage per cycle} \end{array} \right] = e^{\pi/Q_1} - 1. \quad (2)$$

The subscript on Q is used to denote the condition where R is negative. The equation is obtained from the following relationships.

$$\begin{aligned} Q &= \frac{\omega_0 L}{R} = \frac{L}{R/\omega_0} = \frac{1/2 I_m^2 L}{1/2 I_m^2 R/\omega_0} \\ &= \frac{\text{Energy stored in resonant circuit}}{\text{Rate of change of energy stored}} \end{aligned} \quad (3)$$

Since $R = \frac{\omega_0 L}{Q}$, a negative R , $(-R) = -\frac{\omega_0 L}{Q_1}$,

$$e^{\frac{R}{2Lf_0}} \text{ (from equation 1) } = e^{\frac{-(-R)}{2Lf_0}} = e^{\frac{-(-2\pi f_0 L)}{(2Lf_0 Q_1)}} = e^{\frac{\pi}{Q_1}}. \quad (4)$$

I_m = peak value of current flowing in resonant circuit.

The relationship of energy stored to rate of change of energy stored is a convenient relationship for computing the value of Q_1 in this particular case.

The ability of a transistor to satisfy the requirements of the resonant circuit for sustaining oscillation as well as building oscillation will be discussed later.

Energy which is large in comparison to the power taken by the load is stored in the tuned circuit in order to minimize:

(a) The effect on oscillator frequency of small changes in load current phase angle with trunks of different types, connected.

(b) The effect on oscillator frequency of different values of inductance in the output windings of the associated oscillator circuit.

(c) The effect of surges on the transistor that are transmitted over the connected trunk to the oscillator.

(d) The harmonics produced by the method of supplying power to the tuned circuit. This includes the use of an incorrect feedback adjustment.

The Q for the transformers varied from 60 to 85. The stored energy was then 60 to 85 times the energy dissipated per radian at the oscillating frequency for the particular transformer. By adjustment of the ratio of the turns between windings 1-2 and 1-3 a dissipation in the tuned circuits of from 6 to 10 milliwatts was obtained.

In the six-frequency supply three different coil designs are used. One design is used for the two lower frequencies, a second for the two inter-

mediate frequencies and a third for the last two frequencies. Data on these coils are given in Appendix I.

The frequency of oscillation for each coil is determined by the capacitance used with it and is given approximately by the formula:

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{1}{LC}}.$$

Sufficient inductance is used in these coils so that mica condensers can be economically used with them. Mica condensers are used because of their low temperature coefficient.

Leakage reactance and capacitance between windings make the circuit resonant at more than one frequency. To minimize the current fed back to the emitter under such parasitic conditions, the 4-8 winding was placed next to the core with terminal 8 next to terminal 1. This placed an ac ground next to the end of the winding connected to the emitter. As a further precaution against parasitic oscillations the resistance between emitter and transformer is kept high.

TRANSISTOR OPERATION

It is customary to consider the transistor as an amplifier working into a load represented by the tuned circuit. However since the current in the collector and that in the emitter are intimately related during that part of the cycle when power gain is obtained, the collector circuit can be considered equally well as a negative resistance of a value established by the feedback used and the emitter circuit simply as a load. At best either method is only an approximation due to the nonlinearity of the transistor characteristics for large signals. Representation of the transistor as a resistance puts the requirements in terms of values readily obtained from the static characteristics of a transistor. This form of treatment is therefore used.

The regions in which positive or negative resistance is obtained is illustrated in Fig. 3. The characteristics shown are those for an ideal transistor. That is, the ratio of an incremental change in collector current to an incremental change in emitter current, α , with a constant collector voltage is a uniform value in region 2. Also, that in regions 1 and 3 the slope of the lines in each region is constant. That this is not very different from that obtained from some transistors can be seen by comparison with the actual characteristics shown in Fig. 4.

The division line between regions 1 and 2 represents the magnitude of the dc voltage applied between collector and base, $V_c - V_b$. On the left hand side of Fig. 3 the phase relationship and magnitude of ac voltage

applied to the collector and the emitter current are shown. The condition which limit the ac voltage on the collector to this value will be discussed later.

In region 1 the net voltage on the emitter to base circuit is negative and it has no effect on the current which flows in the collector. The current which flows in the collector will therefore depend purely on the reverse resistance of the collector acting as a diode and the voltage applied between collector and base. This is a positive resistance but it is not always linear. Since the only voltage active in this region is that obtained from the tank circuit, this resistance acts as a load on the tank circuit.

In region 2 the voltage applied to the emitter circuit is positive and the current that flows as a result of this voltage exercises control over the collector current. The external resistance that is used in the emitter circuit is sufficiently high so that the small changes that occur in the emitter input resistance as the emitter current is varied are insignificant. The current that flows in the emitter circuit can therefore be considered vary linearly with the ac voltage. Since for each incremental change in collector voltage a proportional change in emitter current will be obtained, a plot of this relationship for a given ratio of emitter to collector ac voltage will result in a straight line. The slope will be determined by the amount of feedback (emitter voltage). Since a decrease in the absolute voltage on the collector results in an increase in collector current, the line has a negative slope. It therefore represents a negative resistance.

In region 3 the emitter current exercises very little control over the collector current.

A negative resistance represented by $-R_2$ is therefore obtained from 0 to π of the ac wave. If a second transistor were added of identical charac-

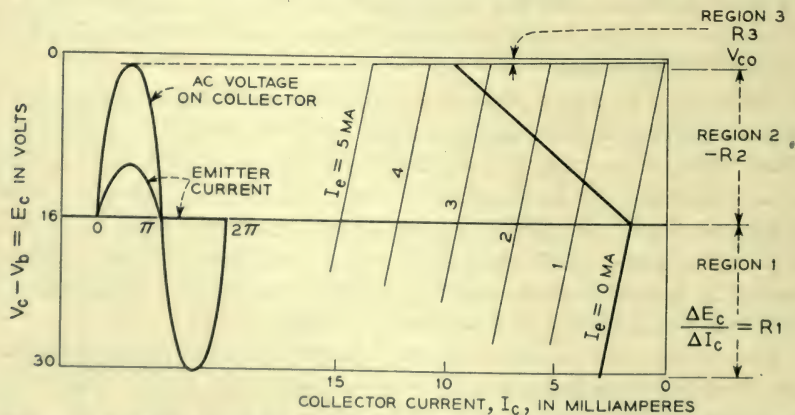


Fig. 3. — Idealized transistor characteristics with operating regions for oscillator indicated.

teristics so connected (push-pull circuit) that a negative resistance would be obtained from π to 2π of the ac wave, the effect of a negative resistance of constant value would be obtained. In order to reduce the transistors to the terms of a two-pole device having a negative resistance the positive resistances represented by the emitter circuits and regions 1 of the transistors would also have to be included.

It is more convenient however to combine all elements that produce a loss when using the static characteristics of a transistor to determine if the transistor will satisfy the circuit requirements for oscillation. This method is therefore used when considering the circuit operation.

There is one important difference between the characteristics of a transistor and the ideal characteristics shown. That is, for very small values of emitter current bias and a constant E_c , the ratio $\Delta I_c/\Delta I_e$, or, α , drops rapidly from three or more to approximately one as the emitter current approaches zero. To eliminate this change in the dynamic negative resistance for very low voltage changes on the collector, a small dc current is supplied in the emitter circuit. The need for this will be discussed later.

CIRCUIT OPERATION

The negative resistance of the transistor is effectively connected in parallel with the resistances representing the various elements that make up the total loss by transformer action. The requirements for oscillation are therefore met when $-R_2$ for a complete cycle is less than the positive resistance representing all losses. If the length of time that $-R_2$ is effective is only one half cycle the value of $-R_2$ must of course be cut in half.* Since a single transistor could meet the requirement for $-R_2$ only one was used in the working circuit as is shown in Fig. 2.

The transistor adopted for this use was the 1729 type, now in production, which has been given the RTMA designation 2N25. The 1729 type was used because its characteristics are least affected by changes in temperature, and in addition the allowable power dissipation was approximately twice that of other comparable transistors.

The various factors that when combined make up the load and their normal variation are given in Table I. All losses are in terms of power into the 1-2 winding.

The corresponding value of load resistance is,

$$R_{TL} = \frac{(V_c - V_b - V_{co})^2 1000}{2 \times 2 \times W_t}$$

* See Appendix II.

TABLE I

Source of Loss	Power in Milliwatts		
	Min.	Avg.	Max.
Sustaining stored energy in resonant circuit.....	6.0	8.0	10.0
Output load.....	4.5	4.5	4.5
Region 1 of collector operation.....	2.8	7.0	14.0
Loss in emitter.....	4.0	8.5	13.4
Margin for stability of adjustment.....	12.0	12.0	12.0
Total, W_t	29.3	40.0	53.9

This would be 1920-ohms, 1400-ohms and 1045-ohms respectively. For purpose of illustration the average value is plotted on the characteristics of an average transistor in Fig. 4 along with $-R_2$ for the transistor. It is evident from this that $-R_2$ is lower in value hence the circuit will oscillate and build up to the required voltage. Minor corrections in the emitter current would normally be required in order to meet test requirements. The potentiometer that is shown in Fig. 2 provides the means for adjustment.

The oscillogram shown in Fig. 5 illustrates the condition described above. The characteristics of the transistor used are shown in Fig. 4. The oscillogram is a multiple exposure from which R_1 , $-R_2$ and R_3 (see Fig. 3) for several values of feedback may be obtained. The normal condition of adjustment is illustrated in Fig. 6 with normal load. Four times normal load is a test condition.

When the extra load that is applied during test is removed the output voltage should go up since power input exceeds the power expended.

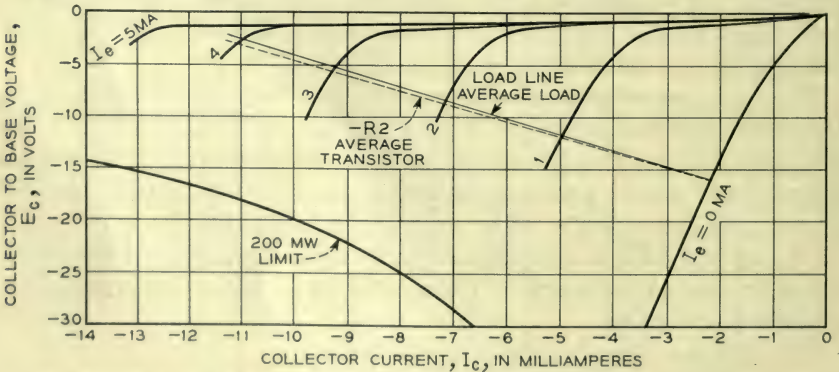
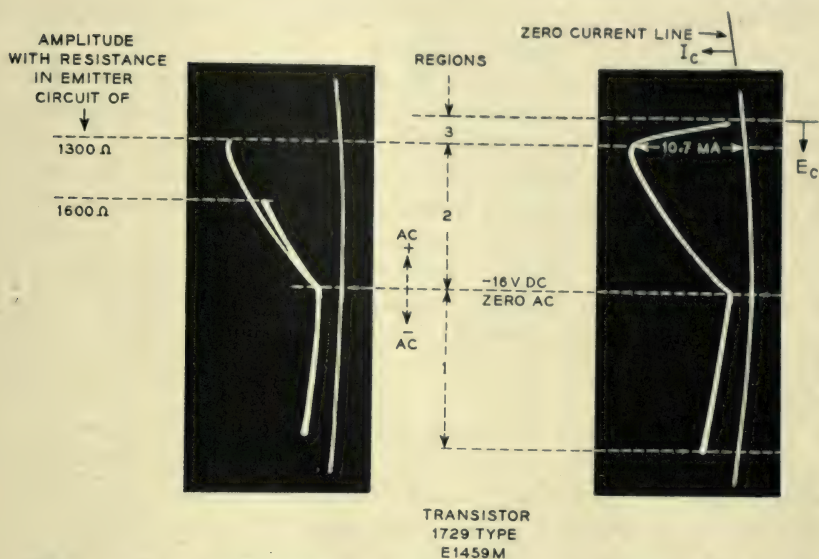


Fig. 4 — Characteristics of representative 1729 type transistor with negative resistance values plotted for the average condition given in table. A load line is also shown for the average condition. In plotting $-R_2$ a line is drawn from $I_e = 0$, $E_c = 16$ to point corresponding to maximum I_e . Maximum $I_e = W_e \times \frac{1.414 \times 2}{E_{rms}}$ at $E_c = 1.5$ volts. W_e is loss in emitter circuit given in table .



Figs. 5 and 6 — Oscillograms showing the relationship of collector current to collector voltage during a complete cycle for several operating conditions. Fig. 5 (left) is a multiple exposure made to illustrate the ability to oscillate at a lower output level with decreased feedback. This is due to the much higher alpha obtained with low emitter currents. Fig. 6 is for normal operating conditions.

This increase in voltage is however very small since the losses will increase as the square of the voltage and the rate at which energy is supplied to the circuit will decrease. The decrease in rate is due to the change from a negative resistance to a positive resistance in region 3. This causes the average negative resistance for the complete cycle to have a higher value.

Feedback which is far greater than is required can in some cases cause the peak value of the ac voltage to exceed the dc voltage. Power is drawn from the energy stored in the tank circuit when this occurs. This effectively limits any further increase in output voltage.

A change will also be introduced in the emitter circuit due to operation in region 3. That is, the voltage feedback introduced in the emitter circuit by collector current flowing in the common base resistance is reversed in phase in region 3.³ This is due to a reduction in collector current when the voltage applied to the emitter circuit is still rising. This feedback is sufficient in many cases to cancel the increase in emitter voltage. The emitter current in such a transistor will therefore remain nearly constant in this region. In region 2 however the feedback is in such a direction as to aid the flow of emitter current. The result is that the voltage drop across the emitter resistance is approximately canceled by the voltage across the base resistance. Due to this relationship the tran-

sistor may appear to have zero resistance in the emitter to base circuit or a small positive or negative resistance.

It is evident from the foregoing that the ability of a transistor to fulfill the requirements for oscillation is therefore dependent upon both the average α along the load line and, R_1 , resistance in region 1. The relationship of these two factors is shown in Fig. 7. Dots on this chart represent the 1729 type transistors tested that met the requirements for the 2N25 transistor. The ambient temperature was $+135^\circ\text{F}$. It is inadvisable to use transistors having a resistance of much less than 4000-ohms in region 1 since both the average dissipation rate and the peak dissipation rate would exceed allowable limits for continuous operation. This will tend to cause the transistor characteristics to change at a more rapid rate.

The effect of different values of feedback on the output of the oscillator is shown in Fig. 8. Fig. 8 also shows the variation in output obtained with several different values of load resistance. This was done to illustrate the use of increased load in determining the proper point for adjusting the feedback resistance. The proper adjustment is the minimum feedback with which the output changes only approximately 1 db in going from normal load to four times normal load. This degree of margin

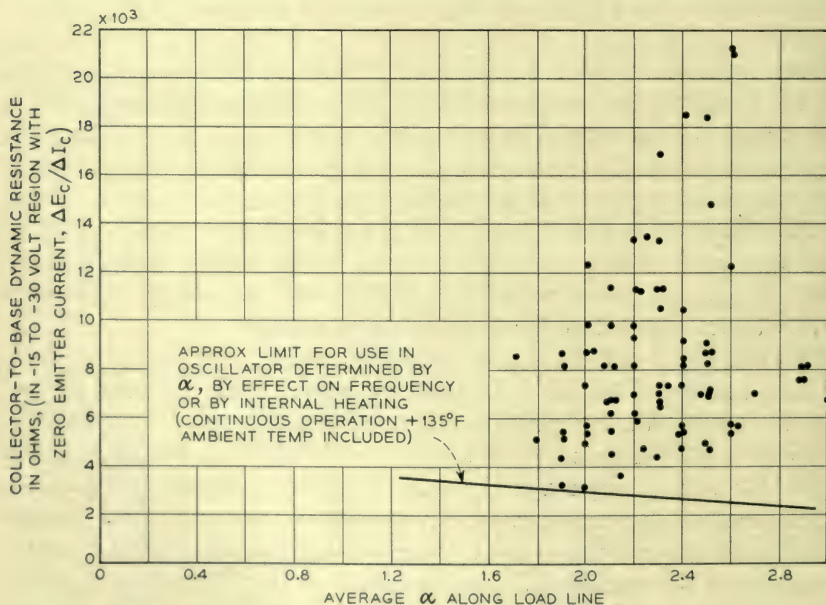


Fig. 7 — Plot of α versus collector to base resistance for representative group of transistors meeting 2N25 transistor requirements.

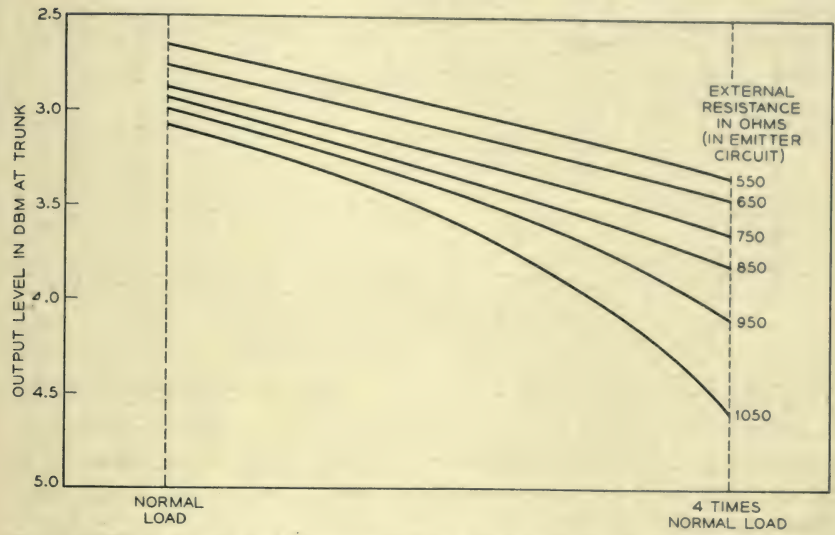


Fig. 8 — Effect of load and value of feedback resistance on the level of the output.

permits some deterioration in transistor characteristics before the change in output is sufficient to require readjustments.

Variation in the absolute level of output due to variations in V_{cc} (see Fig. 3) between transistors and to differences in coils is taken care of by the use of taps on the output windings.

Harmonics of the fundamental frequency are created by the non-linearity of the transistor characteristics. These harmonics are accentuated by excessive feedback. The level of the harmonics for a representative transistor are shown in Fig. 9.

The effect of variations in feedback on the frequency is shown in Fig. 10. The shift is thought to be due to several factors all small. One is the lack of perfect coupling between the transformer windings. Another is

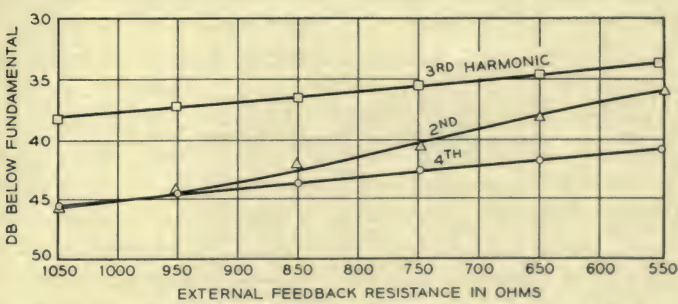


Fig. 9 — Effect of feedback on harmonics with normal load.

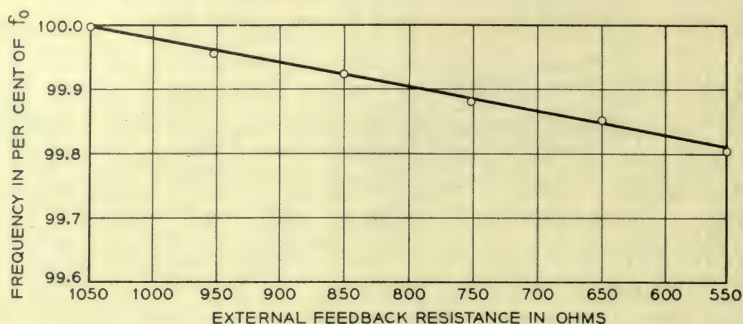


Fig. 10 — Effect of feedback on frequency.

the leakage reactance of the input winding. As the time rate of change in current is increased by increased feedback these factors become increasingly greater although never very large. Fewer turns are required on the coils used for operation at the higher frequencies hence these effects are reduced.

The output level of the oscillator will vary almost directly with the variations in the dc voltage since the amplitude of ac voltage across collector to base is almost equal to the dc voltage applied. Hence, a variation of approximately 0.9 db will be obtained in the output when the central office battery is reduced from 50 volts to 45 volts due to power failure conditions.

The over-all output variation from all causes is shown in Fig. 11. This is based on data obtained using the transistors having the distribution in characteristics shown in Fig. 7.

Decreases in the value of R_1 with temperature is normally compensated by a corresponding increase in α . However small positive or negative voltage changes that alter the level of output do occur in the cut-off voltage. This is minimized by keeping the dc voltage as high as permissible and still meet the 200 milliwatt dissipation limit for the 2N25 transistor.

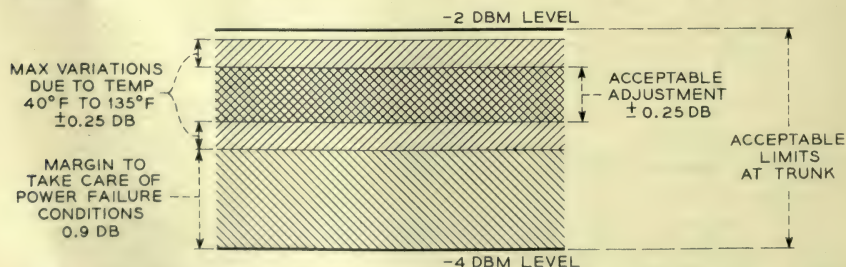


Fig. 11 — Output level with effect of various factors that may alter the level indicated.

The effect of the various factors mentioned before on the frequency of operation are shown in Fig. 12. Since several of the factors causing a shift in frequency were in the negative direction only, the adjustment limits were set correspondingly higher. The over-all frequency variations could be reduced by reducing adjustment tolerances.

The starting condition is important in this type of circuit since energy must be introduced into the oscillatory circuit before the dynamic characteristics of the circuit become effective. This means that the build up time is dependent upon the amount of energy introduced into the system at the start. In this application energy is introduced by the current which flows when the dc voltage is applied to the collector circuit. The value of this current is largely dependent upon the collector to base dc resistance

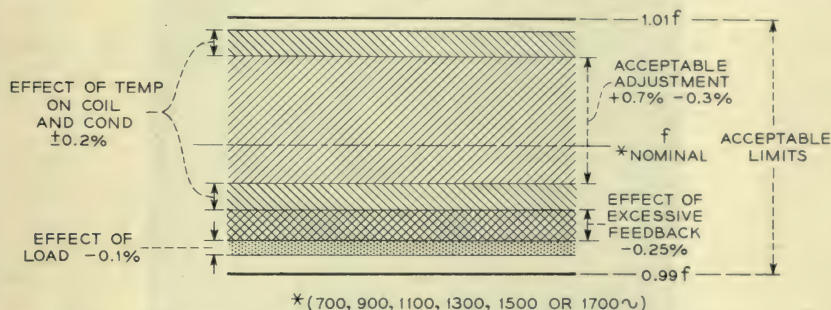


Fig. 12 — Frequency of output with effect of various factors that may alter the frequency indicated.

which is in turn affected by the ambient temperature. This resistance is between approximately 4,000 and 20,000 ohms.

The oscillograms shown in Fig. 13 illustrate this effect. Both are for the same circuit operating under normal conditions of adjustment. Q_1 for this condition is approximately 18. Oscillogram (a) is for the application of voltage in the normal fashion to the voltage divider circuit. The closure occurs at the point oscillation starts. Oscillogram (b) shows the build up obtained when no impulse is applied to start oscillation except for minor irregularities in the dc voltage applied. Starting is prevented in this case by a short circuit on 4-5 winding that was removed approximately 2 ms after the start of the trace. The amplitude of oscillation is so low however for the first few cycles that the start is difficult to distinguish. The build up time for (a) is approximately 27 milliseconds and for (b) it is approximately 37 milliseconds. The exponential build up of amplitude is modified greatly by the rapid change in the transistor's α as the voltage approaches the cutoff region.

It should be noted also that oscillation would not have started under

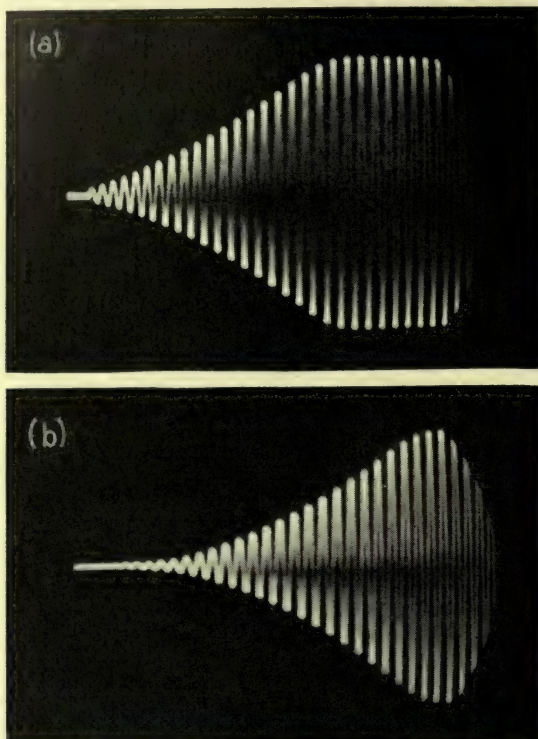


Fig. 13 — Build-up of oscillation in resonant circuit. Normal operating conditions. the condition for oscillogram (b) with some transistors if the small emitter bias current had not been provided.

TRANSIENT EFFECTS

The trunk conductors are balanced with respect to ground. Voltages set up in these conductors due to electrostatic or magnetic coupling to the source of the interference will cause longitudinal currents to flow. An electrostatic shield in the transformer, shown in Fig. 1, effectively prevents such longitudinal currents from reaching the oscillator circuit.

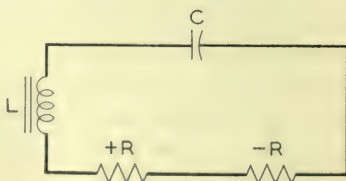


Fig. 14 — Equivalent resonant circuit.

If however the voltage becomes sufficient to breakdown one of the protector blocks that are connected between each trunk conductor and ground, a voltage comparable to the breakdown potential of the protector blocks (400 to 600 volts) would then be impressed across the output windings of the transformer. The usual cause of this condition is lightning. However, neither artificially simulated lightning nor transients of longer duration were capable of raising the voltage on the tank circuit to the point where a transistor was damaged. This is due both to

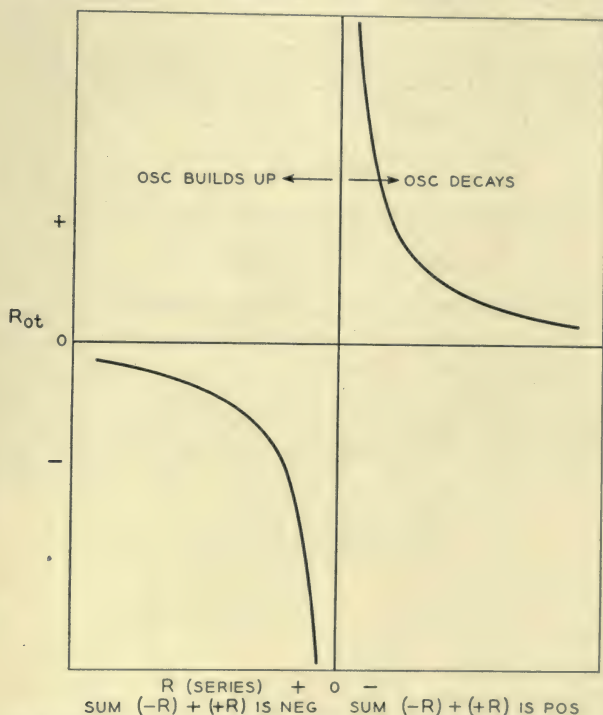


Fig. 15 — Relationship of shunt to series resistance.

the high energy level of the tank circuit and the isolation furnished between repeating coil and oscillator by the series resistors.

SUMMARY:

The transistor oscillator adequately fulfills the requirements for a source of current for multifrequency pulsing over telephone transmission circuits. Adjustments are provided so that the requirements for frequency stability, harmonic level and output level can be met with transistors having a wide range of characteristics. Sufficient margin is provided by

design and by initial adjustment so that an appreciable change in transistor characteristics can be tolerated before readjustment is required.

ACKNOWLEDGMENT

The transformer used in the oscillator was designed by H. E. Vaiden and A. M. King. The transistor used was developed by R. J. Kircher and N. J. Herbert. D. J. Houck assisted in testing the circuit.

REFERENCES

1. Lepage Seely, *General Network Analysis*, McGraw Hill.
2. Kurtz and Corcoran, *Introduction to Electric Transients*, John Wiley and Sons.
3. R. M. Ryder and R. J. Kircher, Some Circuit Aspects of the Transistor, *B.S.T.J.*, **28**, pp. 367-401, July, 1949.

APPENDIX I

Transformers moly-permalloy dust core 1.57 O.D.

700- and 900-cycle operation

Winding (1-2) = 500 turns	Q at 700-cycles, 60
Winding (4-5) = 107 turns	Q at 900-cycles, 68
Winding (4-6) = 115 turns	
Winding (4-7) = 123 turns	
Winding (4-8) = 220 turns	
Winding (1-3) = approximately 5560 (adjusted to meet inductance requirements of $5.2H \pm 1$ per cent)	

1100- and 1300-cycle operation

Winding (1-2) = 372 turns	Q at 1100-cycles, 73
Winding (4-5) = 80 turns	Q at 1300-cycles, 75
Winding (4-6) = 86 turns	
Winding (4-7) = 92 turns	
Winding (4-8) = 164 turns	
Winding (1-3) = approximately 4080 (adjusted to meet inductance requirements of $2.8H \pm 1$ per cent)	

1500- and 1700-cycle operation

Winding (1-2) = 305 turns	Q at 1500-cycles, 82
Winding (4-5) = 66 turns	Q at 1700-cycles, 85
Winding (4-6) = 71 turns	
Winding (4-7) = 75 turns	
Winding (4-8) = 135 turns	
Winding (1-3) = approximately 3360 (adjusted to meet inductance requirements of $1.9H \pm 1$ per cent)	

APPENDIX II

The equivalent circuit for the oscillator is shown in Fig. 14. For stable operation from equation (1), since $f_0 = 1/t$,

$$1 - e^{-\frac{(+R)t}{2L}} = e^{-\frac{(-R)t}{2L}} - 1.$$

The negative resistance, $(-R)$, must therefore be equal to the positive resistance $(+R)$. If however $(-R)$ is active for only half the time, $(-R)$ must be equal in magnitude to $2(+R)$ in order to satisfy the requirements for equality. This assumes that boundary effects are negligible. This assumption was borne out by experiment.

The equivalent resistance of the resonant circuit is

$$R_0 \doteq \left(\frac{\omega_0 L}{R} \right)^2 R \doteq K \frac{1}{R},$$

where

$$k = (\omega_0 L)^2.$$

A plot of this relationship when positive and negative resistances are combined, is shown in Fig. 15.

In the actual circuit the negative resistance is connected across the 1-2 winding. The 1-3 winding is in the resonant circuit. The equivalent resistance $(R_{0(1-2)})$ of the resonant circuit across winding 1-2, is determined as follows:

$$R_{0(1-2)} = \frac{(\text{turns, winding 1-2})^2}{(\text{turns, winding 1-3})^2} \times R_0.$$

Ferrites in Microwave Applications

By J. H. ROWEN

(Manuscript received May 26, 1953)

Since Hogan's exposition of the extreme usefulness of the microwave Faraday effect numerous other laboratories have begun investigating propagation through ferrites and have made significant contributions to the art. In view of the tremendous interest which is being accorded this work this paper has been prepared to summarize some of the observations and developments to date. The plane wave theory is reviewed briefly with special attention being given to the mechanisms by which power is absorbed by the ferrite. The plane wave theory is then modified to describe various waveguide effects. Finally experimental procedures and results are presented to illustrate the theory and to provide general information regarding the design of devices employing these effects.*

INTRODUCTION

The ferromagnetic Faraday effect occurs at microwave frequencies as a direct result of the dispersion in permeability which is associated with ferromagnetic resonance. The resonance can be explained most simply by stating that the total magnetization vector of a magnetized ferromagnetic material has associated with it an angular momentum arising from the angular momenta of all of the spinning electrons contributing to the magnetization. Because of this angular momentum (which is directed along the same axis as is the magnetic moment but in the opposite direction) the magnetization vector behaves as a top or gyroscope. If it is displaced from its equilibrium position in a steady magnetic field it will not rotate directly into alignment with the field but will precess about the dc field direction at a frequency determined by the strength of the dc field. In the absence of damping this precession would continue indefinitely, but damping losses are such that the precession will damp out in approximately 10^{-8} sec.

* C. L. Hogan, The Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications — The Microwave Gyrator, B.S.T.J., **31**, pp. 1-31, Jan., 1952.

If an alternating field is applied at right angles to the dc field the magnetization will be driven in precession and when the driving frequency coincides with the natural resonance frequency as determined by the strength of the dc field a large amount of power will be absorbed from the driving field. Off resonance the power absorption is small, but the effective permeability seen by the driving field will go through a dispersion such as is exhibited by all resonant systems. With this model in mind we can proceed to discuss the phenomenon of ferromagnetic resonance and the ferromagnetic Faraday effect.

INFINITE MEDIUM — LONGITUDINAL FIELD

Polder¹ has shown that because of the gyroscopic nature of the magnetization a tensor permeability is required to relate the magnetic flux density and field intensity vectors in a ferromagnetic medium. At low frequencies the off-diagonal components of this tensor are negligible, and the tensor reduces to the ordinary scalar permeability. At frequencies above about 100 mc. these off-diagonal components can become significant depending upon the magnetic state of the material. When this tensor permeability is introduced into Maxwell's equations and a wave equation is derived for propagation in the direction of the applied magnetic field we find that the normal mode solutions to the wave equation are two circularly polarized waves rotating in opposite directions. A solution in terms of linear polarizations is, of course, possible; but the result is more readily interpreted in terms of the circularly polarized waves. Furthermore, the propagation constant for either circular wave contains a simple scalar permeability, instead of the tensor required to describe the medium in general.

Polder's tensor permeability gives the following relations between $\mathbf{b}$ and $\mathbf{h}$ when there is a static magnetic field along the positive z axis*.

$$\begin{aligned} b_x &= \mu_0 \mu h_x - j \mu_0 \kappa h_y \\ b_y &= j \mu_0 \kappa h_x + \mu_0 \mu h_y \\ b_z &= \mu_0 h_z \end{aligned} \quad (1)$$

The quantities

$$\mu = \mu' - j\mu'' \quad (2)$$

$$\kappa = \kappa' - j\kappa'' \quad (3)$$

are complex relative diagonal and off-diagonal components of the tensor permeability.

¹ D. Polder, *Philosophical Mag.*, **40**, p. 99, Jan. 1949.

* Lower case letters are used for RF magnetic quantities.

Equations giving μ and κ in terms of the applied magnetic field and the fundamental atomic constants are given by Hogan.² These equations show that both μ and κ have a resonance at a value of effective internal field given by $\gamma H = 2\pi f$. If H is in ampere turns per meter and f in megacycles the gyromagnetic ratio is $\gamma/2\pi = 3.51 \times 10^{-2}$.

A plane wave travelling in the z direction in an infinite medium described by equations (1) can be resolved into two counter-rotating circularly polarized plane waves having propagation constants as follows:

$$\Gamma_{+} = j\omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon(\mu - \kappa)} \quad \Gamma_{-} = j\omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon(\mu + \kappa)} \quad (4)$$

where the subscripts $\pm$ refer to positive and negative circularly polarized waves.* The terms inside parentheses are effective scalar permeabilities completely describing the medium for a circularly polarized wave. Calculated values of these effective permeabilities are plotted as functions of applied field in Fig. 1 for a ferrite operating at three frequencies.

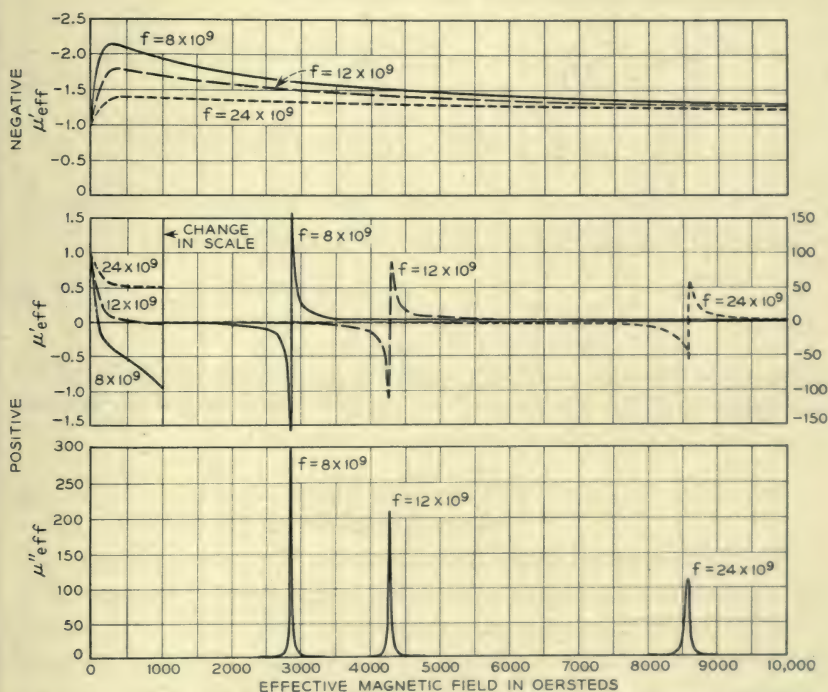


FIG. 1 — Calculated effective permeabilities for positive and negative circularly polarized plane waves computed for three different wave frequencies.

² C. L. Hogan, *Revs. Mod. Phys.*, **25**, Jan., 1953.

* A positive circularly polarized wave is one which rotates in the direction of the positive current producing the dc magnetic field.

From these curves we may calculate the rotation per unit path length, the absorption of the positive component, the net insertion loss and the ellipticity of the resultant wave.

The rotation per unit length is given by

$$\frac{\theta}{l} = \frac{1}{2}(\beta_- - \beta_+) \quad (5)$$

where $\beta_{\pm}$ are the imaginary parts of the propagation constants, $\Gamma_{\pm}$. Let us consider the special case in which the dielectric loss is zero. For convenience we define the complex effective permeabilities seen by the circularly polarized waves as follows:

$$\begin{aligned} \mu_+ &= \mu - \kappa = \mu'_+ - j\mu''_+ \\ \mu_- &= \mu + \kappa = \mu'_- - j\mu''_- \end{aligned} \quad (6)$$

The propagation constants may then be written:

$$\begin{aligned} \Gamma_+ &= \omega\sqrt{\mu_0\epsilon_0} \sqrt{\frac{\epsilon}{2} [\sqrt{|\mu_+| - \mu'_+} + j\sqrt{|\mu_+| + \mu'_+}]} \\ \Gamma_- &= \omega\sqrt{\mu_0\epsilon_0} \sqrt{\frac{\epsilon}{2} [\sqrt{|\mu_-| - \mu'_-} - j\sqrt{|\mu_-| + \mu'_-}]} \end{aligned} \quad (7)$$

It is of particular interest to consider what happens to β_+ when μ_+ becomes zero or negative. If we rewrite the expression for β_+ :

$$\beta_+ = \frac{\omega}{c} \sqrt{\frac{\epsilon}{2}} \sqrt{\sqrt{(\mu'^2_+ + \mu''^2_+)} + \mu'_+} \quad (8)$$

we see that, when μ'_+ is zero or negative, β_+ depends primarily on the magnitude of μ'' , for wherever μ'' is negligible, β_+ is zero. Furthermore, we see that the attenuation constant, α_+ , given by:

$$\alpha_+ = \frac{\omega}{c} \sqrt{\frac{\epsilon}{2}} \sqrt{\sqrt{\mu'^2_+ + \mu''^2_+} - \mu'_+} \quad (9)$$

becomes dependent primarily upon μ'_+ when μ'_+ becomes negative so that we observe a significant attenuation long before μ''_+ becomes large. In Fig. 2 are shown the rotation of the plane of polarization of the linearly polarized wave and the absorption of the positive circularly polarized component of the wave. The dielectric constant of the ferrite was assumed to be 9.0, a typical value for many ferrites.

From these curves it is evident that the wave will be elliptically polarized whenever the effective field is large enough to make μ'_+ zero or

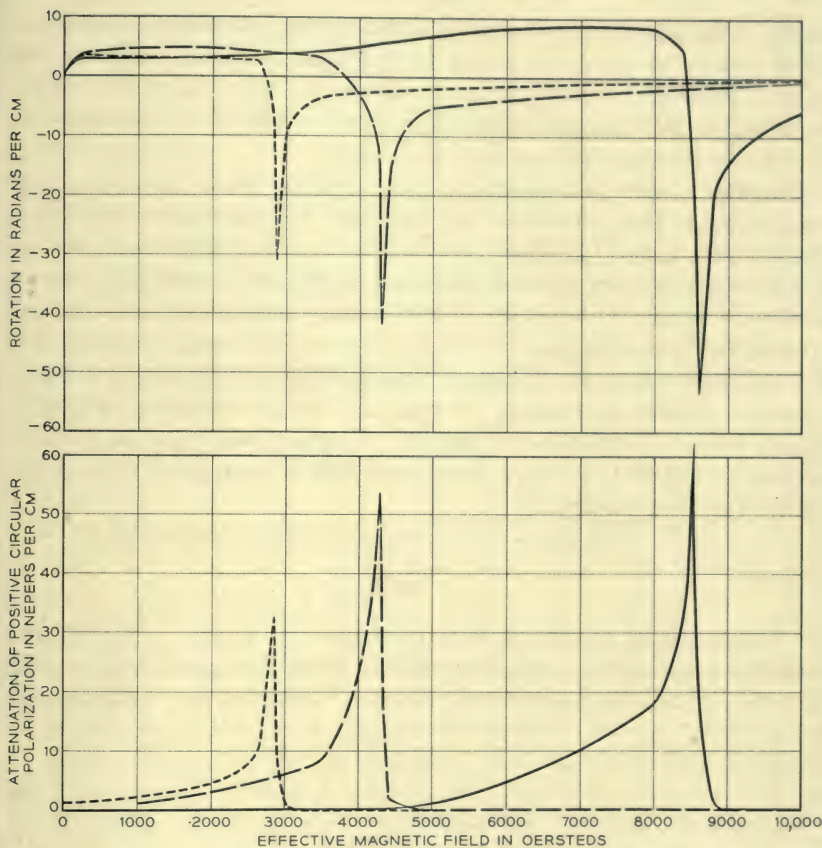


FIG. 2 — Rotation of the linearly polarized plane wave and absorption of positive circularly polarized component versus effective static field computed from data of Fig. 1.

negative, the amount of ellipticity depending in part upon the distance the wave has traveled in the ferrite medium.

PLANE WAVE, TRANSVERSE FIELD

If a wave is propagated in either the x or y direction when the dc field is in the z direction then the wave equation has two orthogonal solutions representing linearly polarized waves. One of these is polarized with the electric vector parallel to the applied dc field and the other has the magnetic vector parallel to the applied dc field. When the magnetic

vector of the wave is parallel to the magnetization the torque on the electrons is zero and the wave sees an isotropic dielectric medium with relative permeability equal to unity. However, when the electric vector is parallel to the magnetization the magnetic vector is at right angles to it and can set the electrons into precession.

Consider a wave propagating in the x direction in an infinite medium magnetized in the z direction. Let this wave be polarized so that it has components E_z and h_y . When this wave enters the magnetized medium, h_y exerts a torque on the magnetization vector M causing it to precess about the z axis. This results in both an m_y and an m_x component of alternating magnetization. There is, however, no component of $\mathbf{b}$ in the x direction because of internal demagnetizing fields arising from an effective volume distribution of magnetic charge as shown in Fig. 3. Such a volume distribution of magnetic charge arising from the periodic reversal in phase of the driving magnetic field is propagated through the medium at the velocity

$$v = \frac{c}{\sqrt{\mu\epsilon}}$$

An instantaneous picture of this distribution is shown in Fig. 3. The magnetic poles set up a magnetic field in the x direction throughout the material. This field is commonly called a demagnetizing field and for

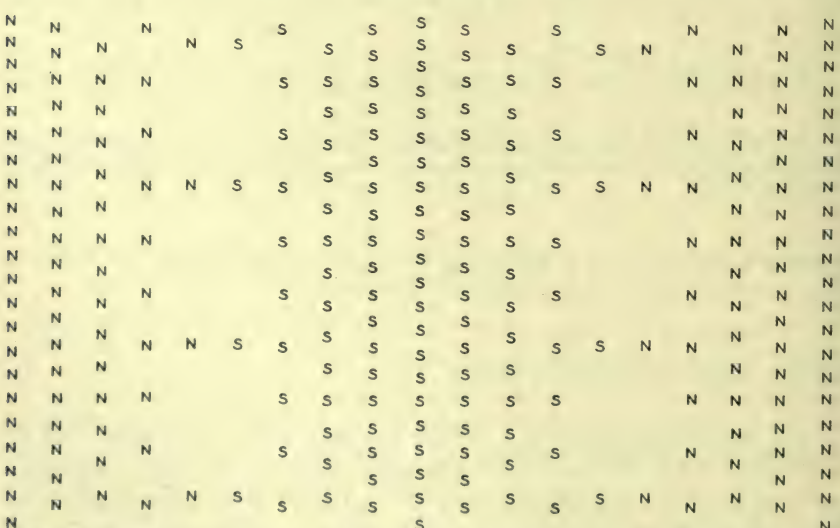


Fig. 3 — Effective volume distribution of magnetic poles arising from phase reversals in transversely magnetized infinite medium.

this particular case it can easily be shown that at every point in the material this field is given by*

$$h_x = -m_x$$

and by definition

$$b_x = \mu_0(h_x + m_x) = 0$$

In one sense this wave is no longer a plane wave as it has a component of $\mathbf{h}$ in the direction of propagation. However, the electric field, E , and the magnetic flux density, $\mathbf{b}$, are unchanged and remain the same as in a normal plane wave.

The solution to the wave equation for the foregoing case yields a propagation constant

$$\Gamma = j\omega\sqrt{\mu_0\epsilon_0} \sqrt{\frac{(\mu^2 - \kappa^2)\epsilon}{\mu}} \quad (10)$$

in which the effective relative permeability of the medium is

$$\mu_{\text{eff}} = \frac{\mu^2 - \kappa^2}{\mu} \quad (11)$$

The real and imaginary parts of this expression are plotted in Fig. 4 for three frequencies.

These curves have the same general shape as those for the *positive* circular component of the wave propagated along the dc field direction, but here they apply to the entire linearly polarized wave. Again we have the possibility of zero or negative permeability. In the region just above resonance the real part of the permeability takes on large values and maintains these even after the absorption curve is nearly zero. This suggests that it is possible to adjust the permeability to equal the dielectric constant of the material so that the medium matches free space perfectly. The medium then can be used as a switch by changing the field from the point where $\mu_{\text{eff}} = 0$ to the point where $\mu_{\text{eff}} = \epsilon$.†

In the region between zero applied field and saturation where the curve levels off, the effective permeability changes almost linearly. In the

* We follow Stratton in making M an H -like quantity rather than B -like.

† In a waveguide μ_{eff} must be adjusted to satisfy the condition that

$$1 = \mu_{\text{eff}} \sqrt{\frac{1 - \left(\frac{\lambda}{\lambda_c}\right)^2}{\mu_{\text{eff}}\epsilon - \left(\frac{\lambda}{\lambda_c}\right)^2}}$$

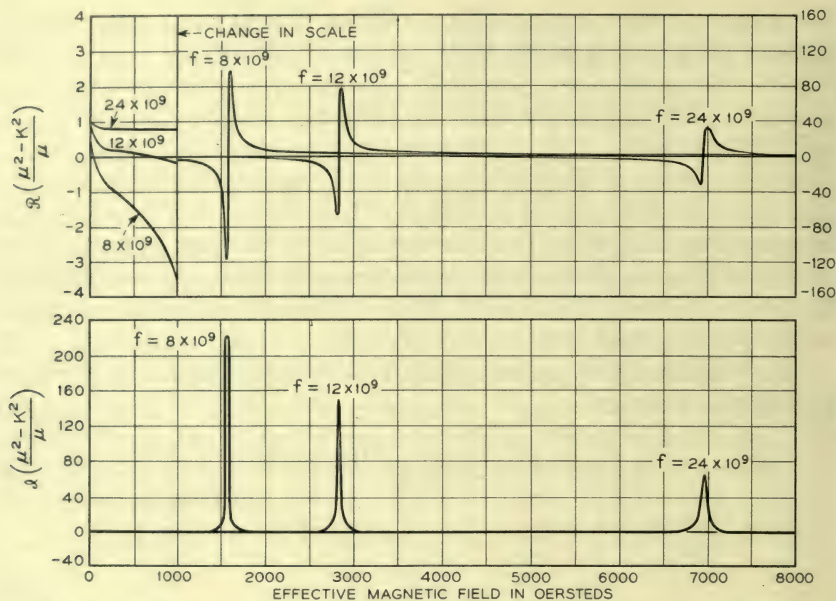


FIG. 4—Effective permeability seen by a plane wave in a transversely magnetized infinite medium. Real part above and imaginary part below.

absence of dielectric loss the propagation constant of the wave is given by:

$$\beta = \omega \sqrt{\mu_0 \epsilon_0} \sqrt{\epsilon} \sqrt{\frac{(\mu^2 - \kappa^2)}{\mu}} = \omega \sqrt{\mu_0 \epsilon_0} \sqrt{\epsilon} \sqrt{\mu_{\text{eff}}} \quad (12)$$

Thus a variable phase delay is obtained by controlling the magnitude of the applied field. This delay is not *necessarily* accompanied by a change in attenuation so that an ideal phase shifter can be made using this effect. There are numerous other applications which can be made of the longitudinal and transverse applied field phenomena. Many of these will be discussed later in this article.

LOSS MECHANISMS IN FERRITES AT MICROWAVE FREQUENCIES

Neglecting dielectric losses, the plane wave theory predicts almost no loss at all for a negative circularly polarized wave and a single absorption line for a positive C.P. wave. In practice a more complicated behavior is observed, and to facilitate the discussion we show in Fig. 5 typical loss characteristics superposed upon the theoretical loss curve of Fig. 1. We will enumerate the main points of interest before proceeding with the discussion. The specific differences in behavior are:

Curve A. Broad resonance absorption line.

Curve B. A loss which disappears when the material is magnetized, called "Low Field Loss".

Curve C. Loss which goes to zero for one component and rises for the other.

Curve D. A loss which appears to be independent of magnetic field over a wide range and can be related to the dielectric loss tangent of the material, hence called dielectric loss.

Curve E. Higher order modes causing erratic variations in loss.

Curve F. Double peaks due to "Cavity Resonances".

Qualitative and semiquantitative explanations have been developed to explain all of these phenomena. Some of them follow from a simple extension of the plane wave theory and the rest are based upon considerations of the special case of a partially filled waveguide.

Curve A, Fig. 5

Associated with the precessional resonance there is a damping term by which power is dissipated in the lattice. The exact nature of this damping term is not fully understood, and measured line widths are always greater than those predicted by present theory. Nevertheless we have at our disposal empirical damping constants which can be used to predict resonance absorption losses as was done in the calculation of the curves of Figs. 1 and 4.

These apply, however, only to small ellipsoidal samples which are ground from single crystal ferrites. In polycrystalline ferrites the absorption line is generally broader for three reasons, namely; crystalline anisotropy, strain anisotropy and varying internal demagnetizing fields due to the variety of shapes of the constituent crystallites.

Many ferrites have a high crystalline anisotropy which behaves in

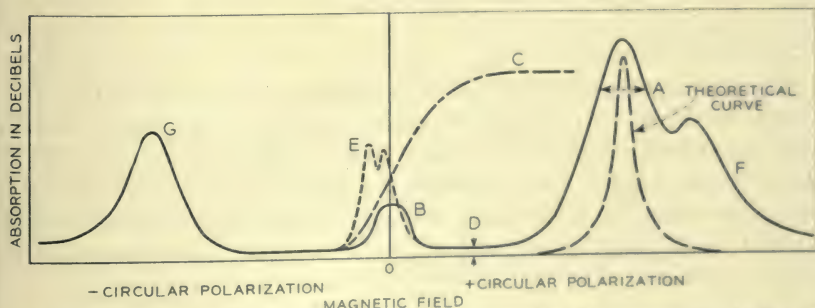


FIG. 5 — Typical loss characteristics encountered in the measurement of various ferrite samples in cylindrical waveguides with longitudinal static field.

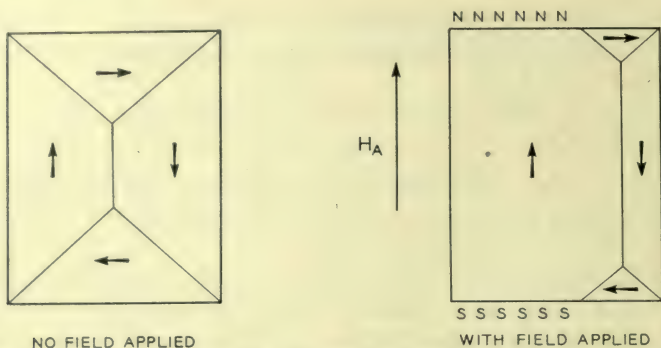


FIG. 6 — A typical domain wall pattern showing the movement of the wall in response to the application of an external magnetic field.

many ways like an internal field tending to keep the magnetization of the constituent crystallites along one of the axes of easy magnetization. In most ferrites there are four such axes and since the magnetization can be in either direction along any one of these, there are eight directions of easy magnetization. When a field is applied the effective internal field is roughly the vector sum of the applied field and the anisotropy field associated with the nearest axis. In a polycrystalline ferrite composed of many randomly oriented crystallites it is evident, therefore, that the internal field varies from point to point in the body so that the resonance absorption line is broadened by an amount proportional to the magnitude of the anisotropy field. A similar broadening can arise from magnetostriction due to fields arising from varying strains throughout the polycrystalline ferrite, and non-uniform internal demagnetizing fields due to the shape of the constituent particles or crystallites can likewise broaden the resonance line. Since a broad resonance line results in ellipticity of the transmitted wave it is desirable to use a ferrite having low anisotropy.

Curve B, Fig. 5

Frequently a loss is observed at low fields as indicated in Fig. 5 by Curve B. Neglecting waveguide effects this hump is symmetrical³ so that it evidently depends on a phenomenon which affects both circular components equally. It is generally agreed that it depends upon the existence of domain walls within the material since it usually disappears as soon as the body is magnetized. However, there is some question as to the specific mechanism involved.

³ Fox and Weiss, *Revs. Mod. Phys.*, **25**, p. 262, Jan., 1953.

A ferromagnetic crystal consists entirely of regions called domains which are completely magnetized along one of the directions of easy magnetization. In general the direction of magnetization of these domains is varied in an orderly manner as shown in Fig. 6 so that the energy of the crystal as a whole is a minimum. In the region between adjacent domains there is a (usually) narrow wall in which the magnetization goes through a gradual change in direction from that of one domain to that of the other. When an external field is applied the magnetization of the crystal is increased by the growth of some domains at the expense of their neighbors. When the crystal is saturated substantially all of the walls have disappeared and the material behaves as a single large domain.

There are currently two proposed mechanisms by which these domain walls could cause a loss at low fields. Becker and Döring⁴ have shown that there can be associated with the motion of a domain wall either relaxation or resonance frequencies. Galt⁵ has measured relaxations in a single crystal of magnetite at 3,000 cps and in a single crystal of nickel ferrite at approximately 2.5 mc and has presented a rather convincing argument that these are due to domain wall motion. In general the *relaxation* frequency would be expected to occur far below the microwave frequencies, but resonances could conceivably occur at microwave frequencies and could be quite broad. Until recently no other theory had been advanced which would explain the losses so often observed at low fields, and these were, therefore, attributed to a high frequency domain wall resonance.

There is a more satisfactory explanation which has recently been stated in different ways by Rado⁶ and by Smit and Polder.⁷ Rado has observed a resonance absorption in the microwave region with zero applied field and has shown from temperature dependence that the frequency of this resonance depends upon the saturation magnetization and the crystalline anisotropy of the ferrite. Smit and Polder have presented a model by which we can see how both of these quantities can enter to produce a loss at low fields.⁷ We consider an ellipsoidal crystallite as shown in Fig. 7. The domain structure shown is one which could exist in some crystallites in a polycrystalline ferrite.

The magnetization in domains numbered 1 will respond to right circular polarization and the others to left circular. In other words a wave rotating clockwise is positive circularly polarized in domains one while a

⁴ Becker and Döring, *Ferromagnetismus*, Springer, Berlin, 1939.

⁵ J. K. Galt, *Phys. Rev.*, **85**, Feb. 15, 1952.

⁶ G. T. Rado, R. W. Wright, et al., *Phys. Rev.*, Nov. 1952.

⁷ D. Polder and J. Smit, *Revs. Mod. Phys.*, **25**, pp. 89-90, Jan. 1953.

wave rotating *counterclockwise* is positive circularly polarized in domains numbered 2. In the absence of any other effects the resonance frequency of all of these domains would be determined by the anisotropy field. However, if we excite both circular polarizations simultaneously and if the relative phase of the two circular polarizations is as shown in Fig. 7(a) poles will be set up at the domain walls as indicated in the figure. The demagnetizing fields associated with these will cause the resonance for both circular components to occur at a frequency given by:

$$f = \gamma H_{\text{eff}} \cong \gamma M_s \tag{13}$$

On the other hand, if the phase of the two circular polarizations is as shown in Fig. 7(b), no poles will be set up on the walls and the resonance will be determined primarily by the anisotropy field.

$$f = \gamma H_{\text{eff}} \cong \gamma H_{\text{anis}} \tag{14}$$

These two examples of the relative phase of the circular waves correspond to linear polarizations in the *x* and *y* directions respectively.

This simplified derivation gives the maximum and minimum frequencies at which resonances can occur. In a material containing a large number of randomly shaped and randomly oriented crystallites, resonances can occur at all frequencies between these limits. Most ferrites have values of *M_s* between 80,000 and 240,000 and anisotropy fields which probably range from 8000 to 80,000 amp. turns/meter with per-

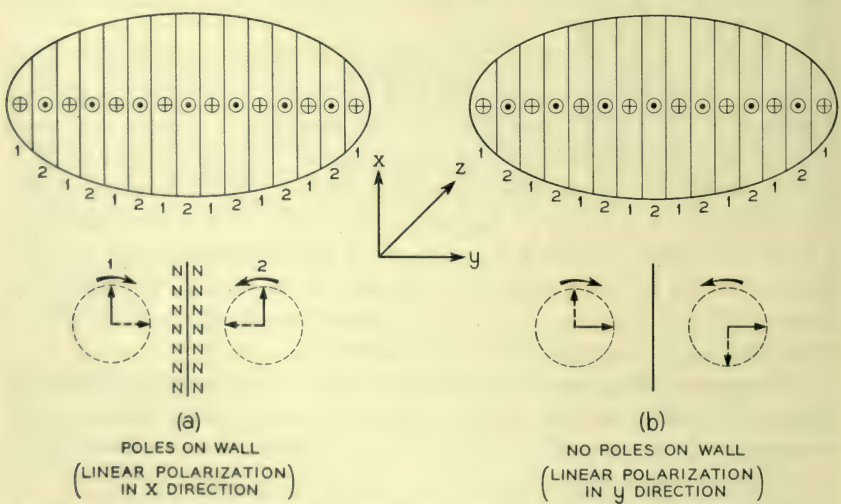


FIG. 7 — Model used by Smit and Polder to illustrate their theory for the "low-field loss."

haps a few which are higher still. Therefore, we have as typical frequencies

$$f_1 = \gamma H_{\text{anis}} \cong 300 \text{ to } 3,000 \text{ mc} \quad (15)$$

and

$$f_2 = \gamma M_s \cong 3,000 \text{ to } 9,000 \text{ mc.} \quad (16)$$

It is evident that at 9,000 mc only that loss associated with M_s will contribute to the "Low Field Loss" and since the mechanism depends upon the existence of domain walls it will disappear when the material is saturated in agreement with our observations.

Curve C, Fig. 5

Some ferrites exhibit a low-field loss which disappears at saturation for only the negative component and increases for the positive component. This is thought to be due to an effective anisotropy field in the material. In order for such behavior to be present at 9,000 mc, however, the internal field must be of the order of 240,000 ampere turns per meter. While such a value of crystalline anisotropy might be found in cobalt or other high anisotropy ferrites it appears to be somewhat too high for a nickel-zinc ferrite such as that in which this characteristic was first observed. However, high internal fields could result from demagnetizing effects similar to those discussed under item B but differing in that the poles are set up on nonmagnetic grain boundaries instead of domain walls. These, of course, would persist when the body is saturated, but there would be loss for only one circular component inasmuch as all of the crystallites are then magnetized in the same direction. Such a loss characteristic can be quite useful where one wishes to absorb one circular component selectively without the necessity for applying a large dc field.

Curve D, Fig. 5

Dielectric losses are present in all of the ferrites which have been made to date, although in some materials this loss is very low. Low dc conductivity in itself is not a sufficient criterion of the dielectric properties of a material as some ferrites appear to consist of conducting regions surrounded by an insulating matrix, and these have fairly high loss tangents at microwave frequencies.

The major mechanism of dielectric loss involves the exchange of electrons between ions in the crystal lattice. It has been found that the presence of ions of the same metal in different valence states on the same

lattice site gives rise to high conductivity and hence high dielectric loss. Conversely, when a ferrite is carefully prepared so that all of the constituents are present in exactly stoichiometric proportion, and when the possibility of multiple valence states is eliminated the conductivity is very low. To illustrate this point a series of measurements is reported in which the iron content of the ferrites was carefully varied about stoichiometry in a nickel-zinc ferrite. These measurements are discussed at a later point in this paper.

A ferrite is made by reacting a mixture of metallic oxides at a temperatures below the melting point of these oxides. As the oxides react a new crystal structure is evolved in which the metallic ions occupy positions in the interstices of a close-packed oxygen lattice. There is a very well authenticated theory due to Neel⁸ explaining the way in which the spin orientations of the ions are distributed in the two types of lattice site which exist in the Spinel oxygen lattice. Whenever metal ions in more than one valence state occupy the same type of site, e.g., the octahedral position, there is a possibility for the easy transfer of an electron from one to the other since the crystal structure is unchanged by the transfer.⁸ In the case of nickel ferrite which has the composition NiOFe_2O_3 an excess of iron will tend to replace some nickel atoms by entering the lattice in the divalent state. Since the remainder of the iron is trivalent, comparatively high conductivity is observed. The problem of producing ferrites with extremely low dielectric losses appears to be fairly well understood and is progressing satisfactorily. By choosing the proper set of metal ions to insure the absence of multiple valence states and by maintaining the proper oxygen stoichiometry one may be able to achieve loss tangents as low as 0.001. The subject of dielectric losses is well covered in the literature.^{9, 10}

Curves E, F and G, Fig. 5

The loss mechanisms indicated in Fig. 5 by Curves E and F all arise from the particular behavior of ferrites in waveguides as differentiated from the plane wave theory.

For example, the erratic behavior indicated by Curve E has been shown to be due to the presence of higher order modes in the ferrite region in a waveguide, and the subsidiary hump on the absorption Curve F has been shown to be a "cavity resonance" which is strongly dependent

⁸ L. Neel, *Physica*, **16**, pp. 350-53, 1950, and *Zeit. Anorg. Chem.*, **262**, pp. 175-184, 1950.

⁹ E. J. W. Verwey and J. H. DeBoer, *Rec. des Travaux Chimiques des Pays-Bas*, **55**, pp. 531-54, 1936.

¹⁰ E. J. W. Verwey et al., *Phillips Res. Reps.*, **5**, pp. 173-187, 1950.

upon the diameter of the ferrite cylinder and upon the guide wavelengths but not upon the length of the cylinder.

Other waveguide effects causing anomalous loss behavior, such as shown by curve G, have been discussed by Fox and Weiss¹¹ and will be treated by them in greater detail in a forthcoming publication.

In order to discuss these effects more fully we must examine the modifications of the plane wave theory which must be made to explain the behavior of a ferrite in a waveguide.

WAVEGUIDE THEORY, LONGITUDINAL FIELD

When a piece of ferrite is placed in a waveguide and magnetized it is necessary to modify the foregoing plane wave theory to describe the behavior of a wave passing through the ferrite. Because of the anisotropic nature of the magnetized ferrite it is necessary to obtain a solution to the specific problem of the waveguide containing the ferrite. When the magnetization of the ferrite precesses about the applied dc field it sets up components of $\mathbf{h}$ which do not exist in any of the classical modes, and unless one can deal with small perturbations the solution becomes quite involved.

The modes which can exist in the ferrite will often resemble the classical modes so that for convenience we will refer to them as modified TE or TM modes. Suhl and Walker¹² have obtained solutions for the case of a cylindrical waveguide completely filled with a magnetized ferrite, and they have shown that the modified dominant TE_{11} mode behaves much like the plane wave in the region of small fields but that the behavior of the TM modes cannot be approximated by a simple extension of the plane wave theory. A waveguide large enough to support the dominant mode when filled with air will, when filled with ferrite, support three or four higher order modes including some of the modified TM modes. In this case, it is possible to have several present at the same time with the result that observations of rotation, loss and ellipticity are almost impossible to interpret. Accordingly, we should reduce the size of the waveguide in the ferrite-filled region, and this involves the creation of discontinuities in the waveguide. This is not always necessary, however, because higher order modes will not always be set up in the ferrite even though the waveguide is large enough to propagate them. If care is taken to avoid geometries which favor a given mode the prob-

¹¹ A. G. Fox and M. T. Weiss, *Revs. Mod. Phys.*, **25**, p. 262, Jan., 1953.

¹² A preliminary report of this work has been published in the form of a Letter to the Editor by H. Suhl and L. R. Walker in *Phys. Rev.*, **86**, p. 122, 1952. A more detailed account is scheduled for publication in the *J. Appl. Phys.*

ability of its occurrence will be greatly reduced. In particular it has been found that a flat-ended cylinder completely filling the waveguide can be introduced into the full-sized waveguide without mode complications, but Fox and Weiss¹³ have shown that putting conical tapers on the ends will favor the establishment of the modified TM_{11} mode. In most applications of the Faraday effect the ferrite element is in the form of a very thin pencil at the center of the waveguide so that the mode problem is greatly simplified, but in order to obtain quantitative fundamental information about ferrites themselves it is often necessary to work with a completely filled waveguide. In such cases considerable care must be taken to insure the validity of the measurements.

One method of making impedance measurements which has been used successfully by H. Suhl is to cut a shallow longitudinal slot in the cylinder of ferrite and to make standing wave measurements directly in the medium. Because the slotted section is filled with ferrite the size of the waveguide can be reduced to insure the presence of a single mode. This is restricted to unmagnetized materials as rotation of the plane of polarization would result in radiation by the slot.

Another suggested procedure is to make a transformer from full-size rectangular waveguide to circular waveguide of diameter equal to $d/\sqrt{\epsilon}$ where d is the diameter of a dominant-mode air filled pipe and ϵ is the relative dielectric constant of the ferrite. This transformer can be treated as a four terminal impedance transformer and its network impedances can be determined by measurement. Impedances measured in the air filled guide will have to be transformed through this network to obtain the true impedances of the ferrite-filled guide, but this can be done if the need for the measurements warrants such effort.

An exact solution of the partially filled waveguide is considerably more difficult to obtain than the solution for a completely filled waveguide. Yet this geometry is the one usually used in most practical applications of the Faraday effect. In the absence of an exact solution one must develop simple physical explanations based upon plane wave theory plus intuition for numerous observed phenomena. A theory for the partially filled longitudinally magnetized waveguide can easily be developed from two simple observations. First, we consider the circular components of the wave separately and observe that each sees an effective scalar permeability which is a weighted average of the permeability of the pencil for that component and that of the surrounding medium, and second we postulate that a small enough pencil will not act as a dielectric rod waveguide and will merely create a small perturbation of

¹³ A. G. Fox and M. T. Weiss, *Revs. Mod. Phys.*, **25**, p. 262, Jan., 1953.

the original mode. When the above assumptions are valid the plane wave theory can be extended easily to explain loaded waveguide behavior. In the plane wave case it was shown that a negative effective permeability results in attenuation of the positive circularly polarized wave. Quite a different result is observed in waveguides containing very small cylinders of ferrite. It appears that if the cylinder is small enough not to act as a dielectric waveguide then the negative permeability inside the rod simply is averaged with the permeability of the surrounding region so that the rotation curve (which depends upon the difference between the square roots of the permeabilities seen by the two circular components of the wave) follows the dispersion curve of the permeability of the positive circularly polarized wave, even following the pattern of the permeability when it is negative.

In Fig. 8 are shown measured curves of the rotation of the wave and the absorption of the positive circularly polarized component of the wave as functions of applied dc field for comparison with Figs. 1 and 2. To amplify our arguments we point out that the propagation constant in a waveguide containing a very small pencil of ferrite is of the form:

$$\beta_{\pm}^2 = \beta_e^2 + A \frac{r_1^2}{r_0^2} \left[\frac{\mu_{\pm} - 1}{\mu_{\pm} + 1} \beta_e^2 - \beta_0^2 \frac{\epsilon - 1}{\epsilon + 1} \right] \quad (17)$$

where β_e is the propagation constant of the empty guide

r_1 is the radius of the ferrite cylinder

r_0 is the radius of the waveguide

ϵ is the dielectric constant of the ferrite

and $\mu_{\pm}$ are the effective complex permeabilities ($\mu \pm k$)

A is a constant $\cong 3.2$

We see that the expression within the brackets is finite for all values of $\mu_{\pm}$ except $\mu_+ = -1 + j0$. Accordingly if the damping parameter (or line width) is large enough to insure that μ_+ can never take on this value, there will always be a cylinder radius, r' , for which the perturbation term is small relative to β_e^2 . However, the cylinder diameter does not have to be very large before the above arguments fail and the rotation and loss behavior become quite different.

The cutoff wavelength of the TE_{11} mode in round guide is

$$\lambda_c = 0.1708 d \sqrt{\mu\epsilon}$$

where d is the diameter in centimeters of the waveguide and μ and ϵ are the effective relative permeability and dielectric constant of the medium contained therein. When a cylinder of ferrite is placed at the center of

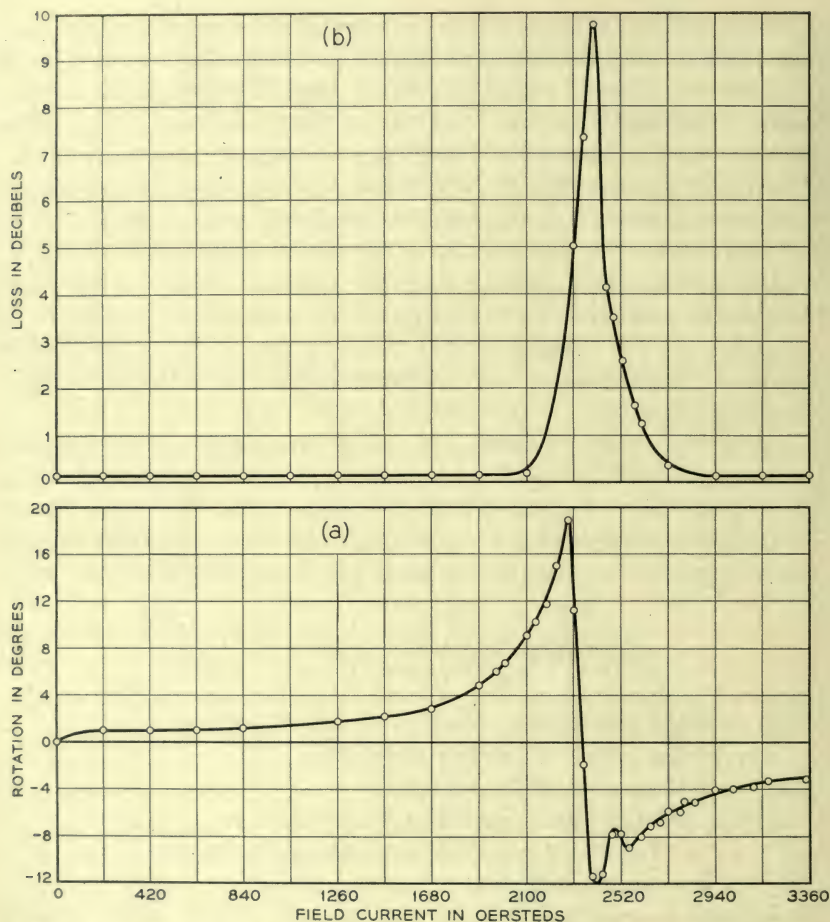


FIG. 8 — Rotation of the plane of polarization and absorption of the positive component as functions of applied dc field for a wave propagating through a waveguide containing a very small cylinder of ferrite. The inflection of 2,400 oersteds is due to a "cavity resonance" as discussed later.

the waveguide the cutoff wavelength is increased because the effective value of ϵ is increased. If the ferrite cylinder is quite small this alteration will be unimportant, but if the diameter of the rod exceeds about one-quarter that of the air filled dominant mode guide this mechanism can lead to the existence of higher order modes for the negative component. When this happens the plane wave theory obviously cannot be expected to apply, for the presence of multiple modes in the propagation of either component will introduce an additional variable not present in the plane

wave theory. Experimentally one will observe very erratic and frequency-dependent behavior under these conditions.

WAVEGUIDE THEORY — TRANSVERSE FIELD

A waveguide, either round or rectangular, filled with ferrite and magnetized by a field parallel to the *electric* vector of the dominant mode will exhibit a behavior qualitatively the same as described in the plane wave theory of the transverse field. In fact, it has been shown that in a rectangular guide all of the TE_{on} modes can exist with only slight modification.¹⁴ That this result is probable may be seen from the fact that the precessing magnetization vector sets up components of $\mathbf{h}$ in the x and y directions when the applied field is in the z direction, and both of these components normally exist in the TE_{on} modes. The primary modification of the mode arises from RF demagnetizing fields in the ferrite. Because of this modification it is extremely difficult to match the boundary conditions for normal incidence at an interface between the ferrite and air in the waveguide. An infinite series of modes is actually required, but in practice the mismatch due to magnetic effects is usually not very large. If one matches the dielectric constant by means of tapered dielectric horns the remaining mismatch is slight except where μ_{off} approaches zero and at resonance.

While the completely filled waveguide magnetized by a transverse magnetic field parallel to the electric vector of the wave will exhibit a reciprocal behavior in respect to phase change and attenuation, an interesting and potentially useful modification of these effects occurs when a small piece of ferrite is located asymmetrically in a waveguide. Chait and Sakiotis¹⁵ of the Naval Research Laboratory and Turner of the Holmdel laboratory of Bell Telephone Laboratories have independently observed a phase shift which is dependent upon the direction of propagation of the wave, and a simple explanation of this effect has been made by Turner¹⁶ and by Kales.¹⁵ Suhl and Kales¹⁵ have shown the theoretical validity of this explanation. The idea can be demonstrated by consideration of the field configuration shown in Fig. 9.

An observer at the point P will see an $\mathbf{h}$ field which is elliptically polarized in a plane normal to the direction of H_A . The sense of the rotation of the larger circular component of the ellipse will depend upon the direction of propagation of the wave. Thus for *one* direction the major

¹⁴ A. A. van Trier, Paper presented orally at meeting of Amer. Phys. Soc., Washington, D. C., April, 1952.

¹⁵ M. L. Kales, H. N. Chait and N. G. Sakiotis, Letter to the Editor, J. Appl. Phys., June, 1953.

¹⁶ E. H. Turner, Letter to the Editor, Proc. I. R. E., **41**, p. 937, 1953.

part will be a *positive* circular polarization with respect to the applied H_A and will experience a decrease in μ' and for the other direction of propagation the $\mathbf{h}$ field will be primarily *negative* circularly polarized and the wave will experience an increase in μ' . Since the $\mathbf{h}$ field is linearly polarized in the transverse plane at the center of the guide and is linear in the longitudinal direction at the edge of the guide it is evident that there is a point in between where the differential phase shift is maximum. Suhl has shown that this point always occurs halfway between the guide wall and the center regardless of the proximity of cut-off.

Obviously one has merely to adjust the length of the sample and the strength of the magnetic field so that the differential phase shift is 180° and he will have a gyrator. Then it is an easy matter to design a circulator, isolator, or any of the numerous devices depending upon the gyrator action.¹⁷

EXPERIMENTAL PROCEDURES AND RESULTS

In order to verify the theory and to determine the optimum performance obtainable in microwave devices employing the Faraday Rotation an extensive measurement program has been set up. Information of both a fundamental and of a practical nature is obtained through a variety of measurements. From a practical point of view we are interested

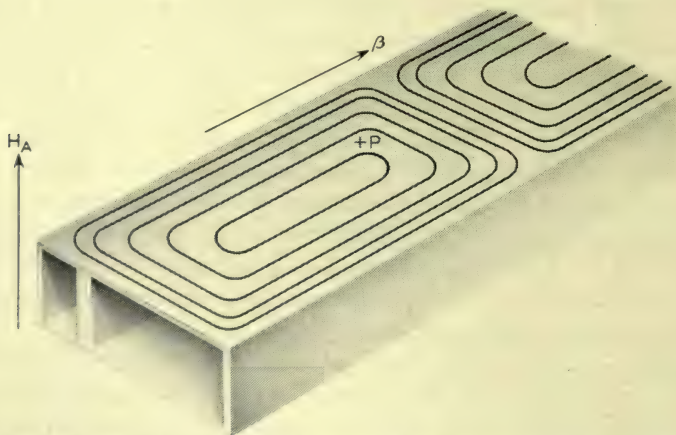


FIG. 9 — Rectangular waveguide containing assymmetrically located ferrite slab. Magnetic lines of force are shown on the top wall of the guide.

¹⁷ C. L. Hogan, The Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications — The Microwave Gyrator, B.S.T.J., **31**, pp. 1-31, Jan., 1952.

in obtaining a ferrite which gives the maximum rotation per unit loss, and at the same time we are interested in obtaining such fundamental information as the mechanisms of loss, the effect of composition, and the relationship between the various physical properties of the ferrite crystals and their microwave behavior. As a corollary interest we wish to relate the microwave behavior of ferrites to their low frequency performance. For convenience most of the microwave measurements have been made in the 9,000 mc X-Band region, but 4,000, 24,000 and 48,000 mc measurements have been carried out by others in Bell Laboratories and some of their findings will be reported here for comparison with those obtained at X-Band.

In the most common procedure the Faraday Rotation and attenuation of a linearly polarized wave and the ellipticity of the resultant wave are measured as a function of the intensity of the applied longitudinal magnetic field. The experimental equipment used in these measurements is shown in Fig. 10(a). A variety of sample shapes ranging from cylinders completely filling the waveguide to very small cylinders suspended along the axis have been measured. From these data the absorption of the positive circular component may be obtained if one assumes that the attenuation of the negative component remains constant as predicted in the plane wave theory. This is not always the case, however, due to the mode configurations sometimes present, and for this reason an alternate

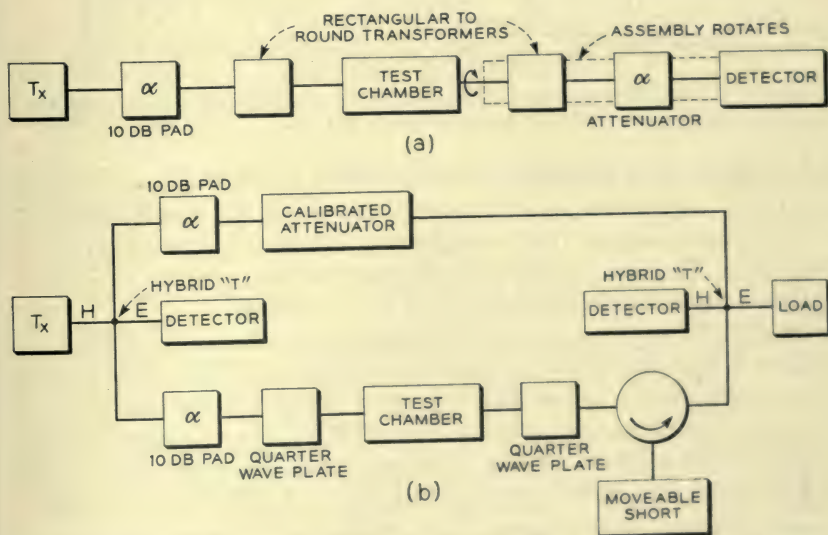


FIG. 10 — Block diagrams of the two most commonly used measuring set-ups.

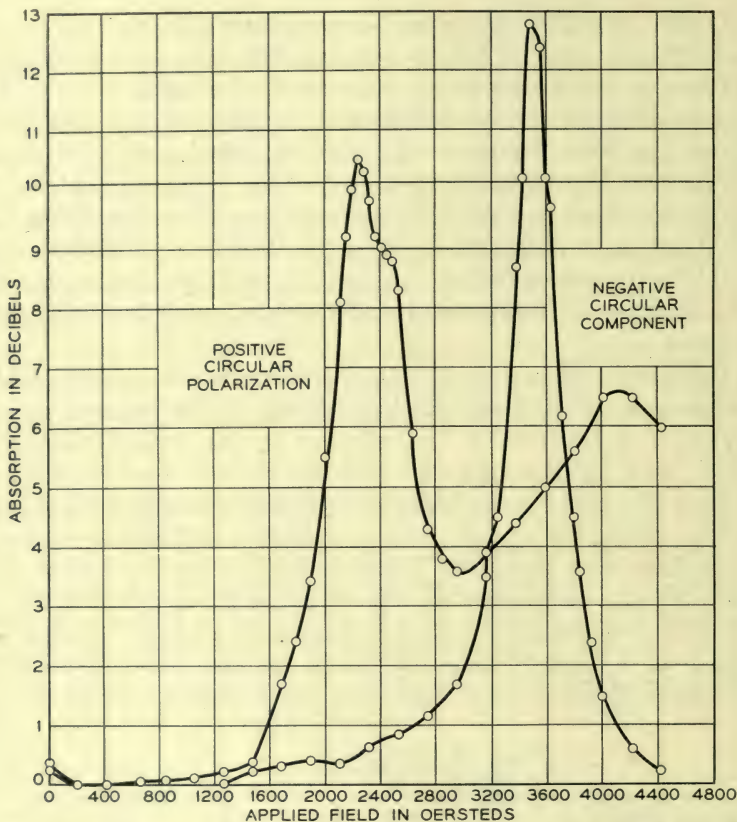


Fig. 11 — Absorption versus applied field for positive and negative circular polarizations. Ferrite cylinder diameter was 0.125".

procedure is often used. This consists simply of exciting the circularly polarized components separately and measuring the phase shift and attenuation of each of these components with the equipment shown in Fig. 10(b). There is no ambiguity in these results, and the equivalent rotation can easily be calculated from the phase data.

Because Polder's permeability matrix is made linear by approximations it is desirable to verify that the measurements made with circular polarization can be superposed to produce the same result as would be obtained using linear polarization.¹⁸ For this purpose we choose a ferrite sample which exhibits a fairly complicated loss curve for both circular components. The absorption characteristics for both circular polariza-

¹⁸ An experiment of this type was first performed by J. P. Schafer of Bell Telephone Laboratories in 1951 to establish the veracity of observed Faraday Rotations.

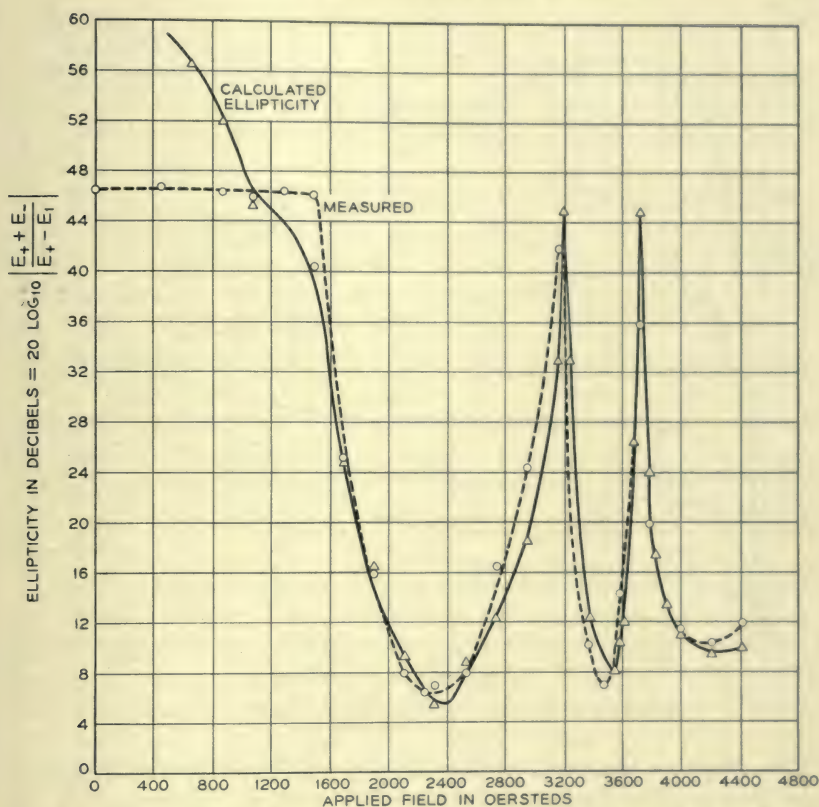


FIG. 12 — Ellipticity versus applied field as observed and as calculated from the data of Fig. 11.

tions are shown in Fig. 11. The resultant ellipticity in db is calculated from the relation

$$\Delta = 20 \log_{10} \left| \frac{E_+ + E_-}{E_+ - E_-} \right| \quad (18)$$

where $E_{\pm}$ are the rms values of the circular components. This result is shown by the solid line in Fig. 12, and the ellipticity as measured by the set-up of Fig. 10(a) is shown by the dashed line. The agreement between these two curves is quite good if we consider that the accuracy of measuring each circular component is no better than ± 0.05 db so that in the regions where the two circular components are actually equal the calculated ellipticity can vary from infinity down to 52 db, while the signal to noise ratio of the equipment limits the measured ellipticity to 50 db.

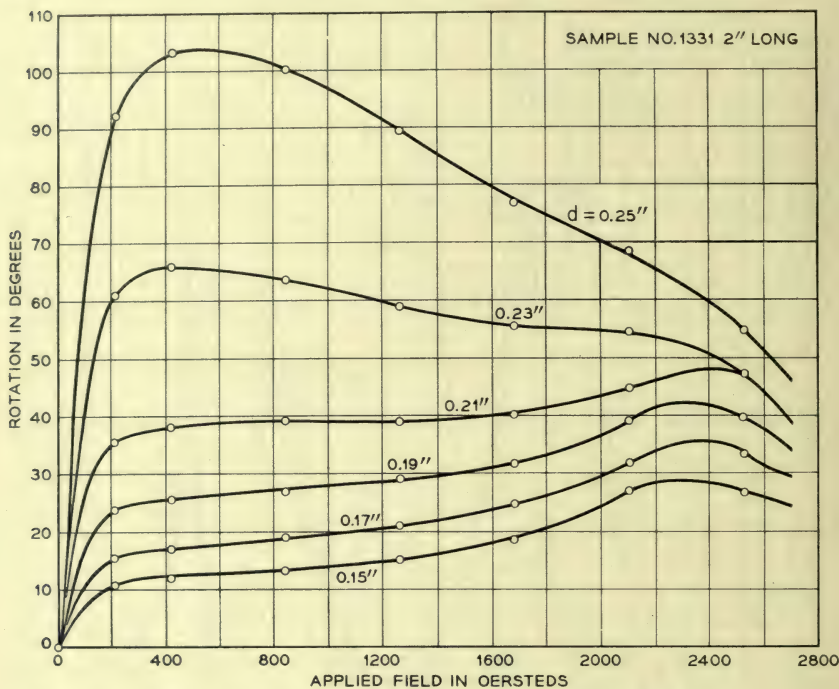


FIG. 13 — Dependence of rotation upon ferrite cylinder diameter.

To illustrate the effect of the diameter of the ferrite specimen upon the Faraday Rotation the following experiment was performed. A one-quarter inch cylinder suitably tapered on both ends was measured in the rotation measuring equipment and then ground down in steps of about 0.020" and remeasured after each cut. The particular ferrite chosen was one having very low dielectric loss and a dependence upon sample diameter which is typical of many ferrites. The data are shown in Fig. 13. At the largest diameter we observe a rotation which agrees with neither the plane wave theory nor the small perturbation approach. This behavior is evidently due to the presence of a higher order mode of propagation in the section of waveguide containing the ferrite cylinder.

Higher order mode effects are undesirable from the standpoint of bandwidth as the rotation produced thereby is much more frequency sensitive than that obtained from the TE_{11} mode. This is evident from consideration of the fact that the partially filled waveguide is much closer to the cutoff for the higher order modes than for the dominant mode, and we will show that frequency dependence is increased as waveguide approaches cut-off for a given mode.

BANDWIDTH OF THE FARADAY ROTATION

Some consideration has been given to means of increasing the bandwidth over which the Faraday Rotation is relatively constant. Two possible ways of broadbanding the effect can be suggested on the basis of measurements and theory.

In the infinite medium-plane wave theory the Faraday Rotation is shown to be totally independent of frequency everywhere far above the ferromagnetic resonance frequency. One way of explaining this lack of dependence upon frequency is to observe that the rotation per unit wavelength *decreases* with increasing frequency while the number of wavelengths per unit length *increases* at the same rate so that the rotation per unit length remains constant. In a waveguide there are two effects which cause the rotation to be frequency dependent. In the first place the guide wavelength is not linearly related to frequency and secondly, where thin pencils of ferrite are used, the radial distribution of field varies with frequency in such a way as to add to the frequency dependence.

By surrounding the ferrite element with a material of the same dielectric constant as that of the ferrite the guide wavelength will be reduced to approximately one-fifth the cut-off wavelength and will be almost linearly related to the frequency in this region. Furthermore, the radial distribution will not change appreciably with frequency because, neglecting the difference in permeabilities, the waveguide is now filled with a uniform dielectric. On this same basis there should be no tendency for the structure to set up higher order modes, so that if care is taken in the design of the transitions from circular to rectangular waveguide, the higher order modes can be avoided.

While the above approach would give very good bandwidth a lack of very low-loss dielectrics having the proper values of dielectric constant limits its usefulness. One of the best dielectrics in this range is Micalex K, and it has a loss tangent of approximately 0.001 which would produce a total loss of several tenths of a db in a length of two or three inches.

The bandwidth can be increased in another way which requires no special dielectric materials. Suhl has shown that when the ferrite element is not perfectly matched to the waveguide the resulting multiple reflections can enhance or detract from the inherent rotation. When the element is an integral number of half wavelengths long the resulting rotation is maximum and when the element is an odd number of quarter-wavelengths long the rotation is minimum. Since the rotation of a perfectly matched cylinder increases with increasing frequency we must choose a length for the unmatched element such that it is an integral

number of half-wavelengths long at the low end of the frequency band in order to increase the rotation at that point. The ferrite chamber may be designed as a resonant cavity very tightly coupled so that over the band the input impedance satisfies the matching requirements even though there are multiple reflections inside the cavity.

Some ellipticity of the resultant wave will be introduced by this method as the wave will be linearly polarized only at those frequencies at which the ferrite element is an integral number of quarter-wavelengths long. However, if only a small correction is applied in this way the ellipticity will be tolerable.

The first and second broadbanding techniques may be combined in a manner which was discovered in the course of our measurements to determine bandwidth. Consider a circular waveguide such that at the low end of the band the wave is just slightly above cut-off. Then the wave impedance in the air-filled pipe will be high. If now the pencil of ferrite is supported in a polystyrene holder, the wave impedance in the ferrite region is much lower than that in the air filled region and multiple reflections will be set up. In order to reduce the magnitude of the reflections to achieve just the right degree of correction each end of the polystyrene can be tapered at the proper angle to give optimum correction. The polystyrene alone gives some improvement in bandwidth according to the arguments first submitted, and the internal reflections provide the rest of the compensation. The data shown in Fig. 14 show the variation of rotation with frequency obtained in this manner. The dotted extension of the low end of the curve indicates the expected rotation in the absence of the correction. By means of this compensation we are able to restrict the variation in rotation to ± 4 per cent over a band wider than 15 per cent. Over this band the ellipticity defined as the ratio in db of the maximum to the minimum field strength was over 50 db which corresponds to perfectly linear polarization within our ability to measure it. As a further advantage of this system the rotation obtained at all frequencies from the ferrite pencil is increased through the use of the polystyrene. This amplification of rotation follows directly from an extension of the plane theory.¹⁹

The rotation per unit length is given by the relation

$$\frac{\theta}{l} = \frac{\omega}{c} \sqrt{\epsilon} (\sqrt{\mu_-} - \sqrt{\mu_+}) \quad (19)$$

where $\mu_{\pm}$ are the effective permeabilities seen by the circularly polarized

¹⁹ Such an amplification of rotation was first observed and studied by J. P. Schafer of Bell Telephone Laboratories at Deal, N. J.

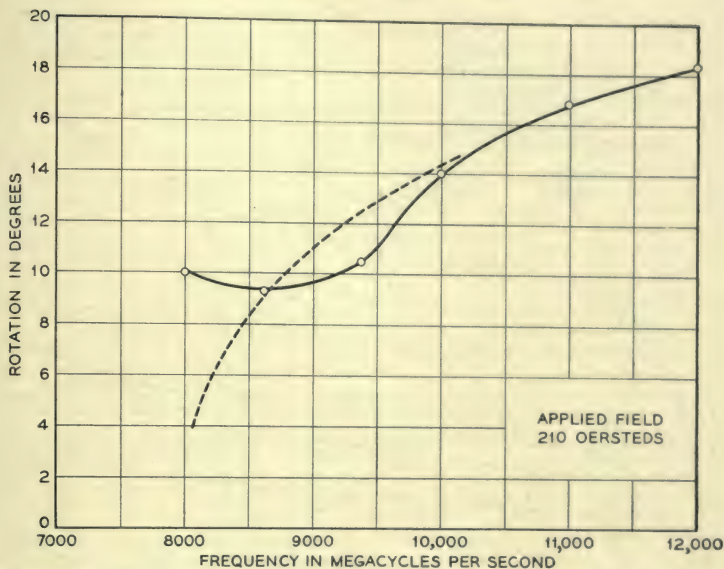


FIG. 14 — Variation of rotation with frequency showing broadbanning obtained through compensation technique.

components of the wave. The corresponding relation for the partially filled waveguide will be a transcendental expression involving $\mu_{\pm}$ and ϵ in a similar way so that arguments regarding the waveguide problem may be based upon equation 15 if we consider $\mu_{\pm}$ and ϵ to be effective values averaged over the guided mode. Thus by increasing the dielectric constant of the region surrounding the ferrite we increase the rotation by increasing the average value of ϵ . Since we also change the radial distribution in such a way as to reduce the fraction of the power contained in the ferrite the amplification in rotation will be less than would be obtained if the waveguide diameter were reduced at the same time. In Fig. 16 we show the effect of increasing the dielectric constant of the region surrounding the ferrite and the effect of a subsequent reduction in guide size. Here again we see that if very high dielectric constants were available in low loss materials a really significant improvement in performance could be obtained. Nevertheless, even the effect of the polystyrene is useful and here we suffer a loss of less than 0.1 db at X-Band.

In the discussion of the loss mechanisms the second hump on absorption Curve F in Fig. 5 was described as a "cavity resonance". While the exact mode of resonance cannot be determined except from a complete solution of the partially filled waveguide problem, we are able to show that the subsidiary hump is strongly dependent upon the diameter of

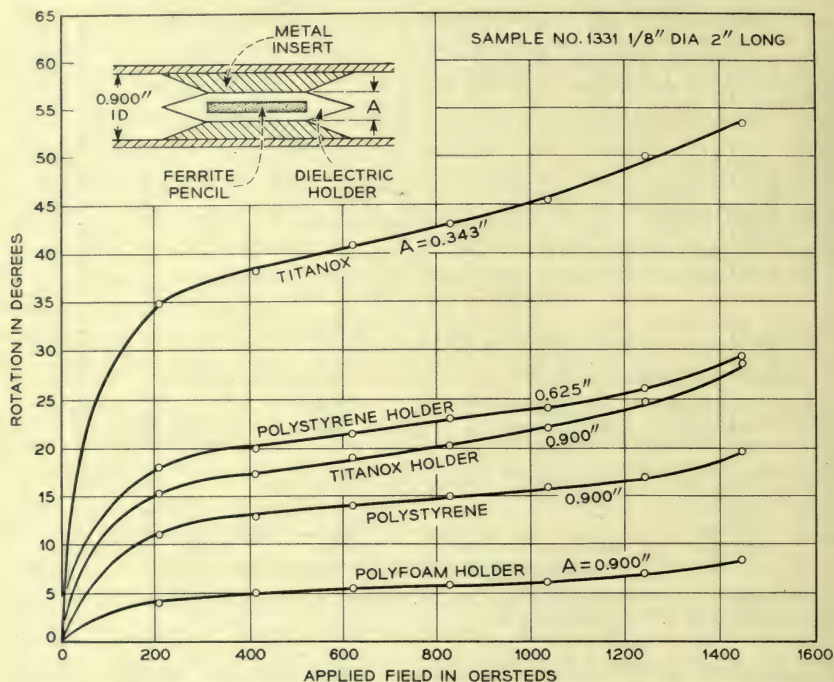


FIG. 15 — The dependence of the rotation characteristic upon the geometry and dielectric constant of the surrounding medium.

the ferrite element and upon the guide wavelength but that it is entirely independent of the total length of the element. In Fig. 16 we show the absorption of a positive circularly polarized wave as a function of applied field for several cylinder diameters. We see that the first hump remains relatively fixed in position while the second hump moves rapidly in the direction of higher fields and becomes larger in height as the diameter is increased. The first resonance is the ferromagnetic resonance, and its movement with increasing diameter is due simply to the fact that the demagnetizing factors of the cylinder are changed when the diameter is changed. The movement of the second hump can be explained if we assume that the resonance frequency of the shape resonance is given by an equation of the form

$$f = \frac{A}{d\sqrt{\mu'_+ \epsilon}} \quad (20)$$

This is an equation typical of those encountered in dimensional resonances. Here A is some constant associated with the geometry of the sample and d is the cylinder diameter. At a fixed frequency, we see that

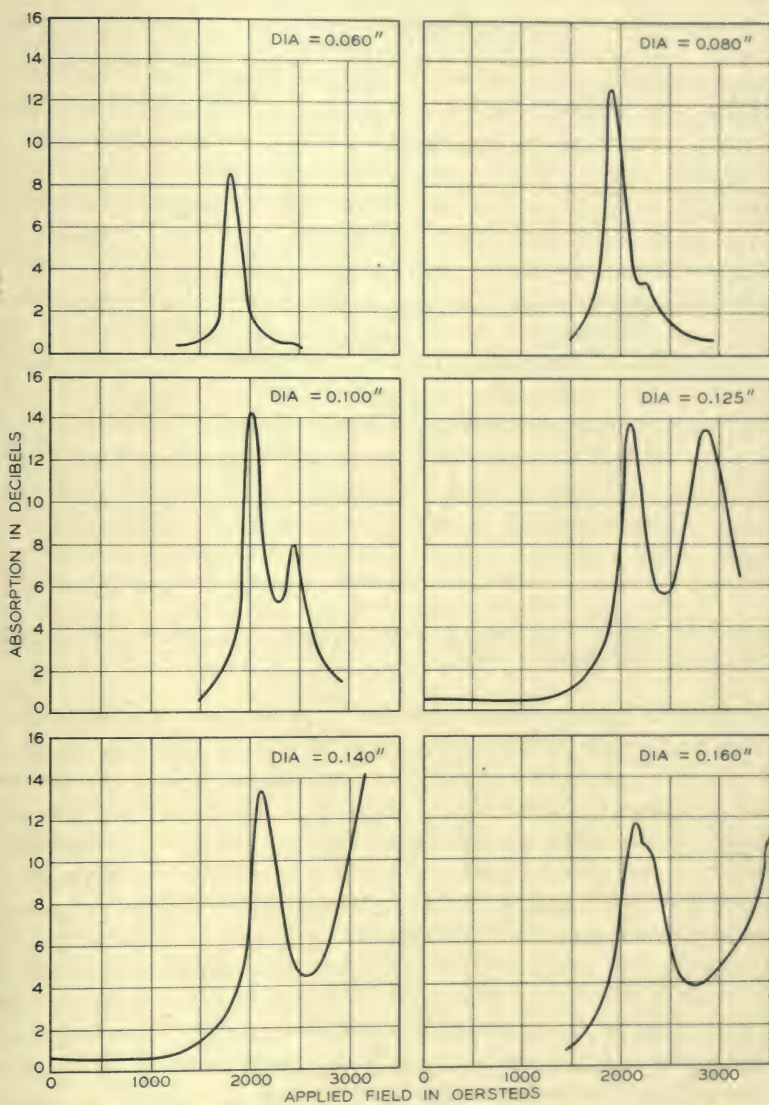


FIG. 16 — Absorption of a positive circularly polarized wave for several sample diameters showing "cavity resonances".

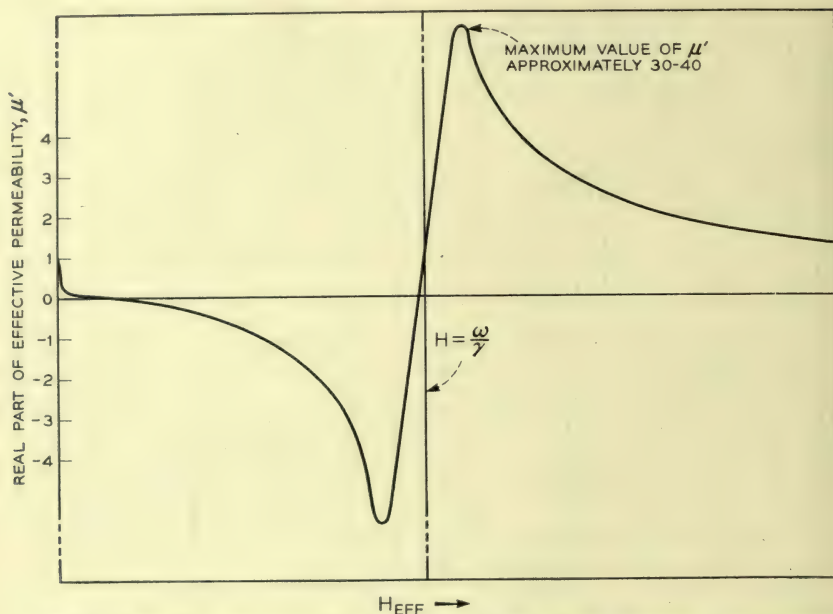


FIG. 17 — Exaggerated scale plot of the real part of the effective permeability versus effective dc magnetic field.

an increase in d must be accompanied by a decrease in μ'_+ to maintain resonance. For convenience we reproduce in Fig. 17 the behavior of the real part of the effective permeability as a function of field. We see that above resonance the permeability starts at a very high value and then decreases as we move toward higher fields. Thus as we increase the diameter, d , we would expect the "cavity" resonance to move to the right, and since in doing so we go further from the ferromagnetic resonance absorption we expect the height of the peak to increase due to the fact that the loss tangent of the permeability is smaller, and hence the effective Q of the "cavity resonance" is larger. Finally we observe that the broadening of the peak which occurs in the larger diameters is due simply to the fact that the slope of the μ'_+ curve decreases as we move to the right so that a larger change in dc field is required to take μ'_+ through a given variation.

A similar movement of the subsidiary hump can be effected through a change in the dielectric constant of the medium surrounding the ferrite element. However, altering the total length of the ferrite in steps of 0.030" until the total change was greater than 0.250" produced no noticeable effect upon either the relative heights or positions of the two peaks.

FERRITE COMPOSITION AND IRON STOICHIOMETRY

There is a continuous program of measurement in which a large number of ferrites from various sources are tested at X-Band and other frequencies. In the early days of the program the most remarkable feature of the results was the tremendous variation in properties even among materials of supposedly the same composition. More recently some general trends have been observed, and it is now possible to correlate the microwave performance of certain types of ferrites with their chemical and structural nature. The nickel-zinc ferrites in particular have been studied extensively, and those having the approximate formula $\text{Ni}_{10.3}\text{Zn}_{0.7}\text{Fe}_2\text{O}_4$ are quite good for microwave applications so long as they have slightly less iron than the above formula requires. A series of measurements illustrating the effect of slight variations in iron content yielded the data shown in Figs. 18 and 19. These ferrites were prepared by F. J. Schnettler of Bell Telephone Laboratories in such a way that their bulk properties correspond very closely to their crystallite properties. The correspondence between the bulk dc conductivity and the RF

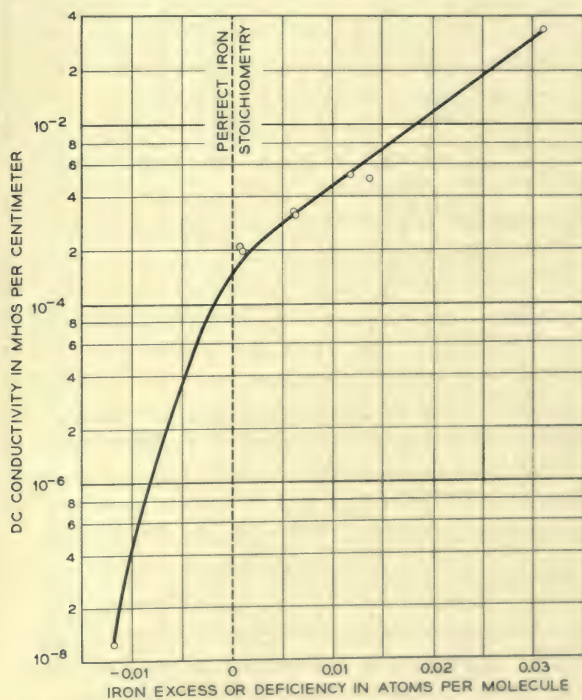


FIG. 18 — DC conductivity versus iron content of NiZn ferrite.

dielectric loss is particularly striking, indicating that the primary mechanism of loss in these ferrites is simply ohmic conductivity loss. This series of measurements clearly shows that the presence of even a very small amount of divalent iron is undesirable in ferrites to be used at microwave frequencies.

TRANSVERSE FIELD MEASUREMENTS

Two classes of measurements are commonly made in which a transverse dc magnetic field is applied parallel to the electric vector of the TE_{01} mode in rectangular waveguide. Those measurements involving a completely filled or symmetrically loaded waveguide yield primarily reciprocal effects which are useful in applications as previously mentioned. More useful, still, are the non-reciprocal effects observed when a piece of ferrite is placed asymmetrically in a waveguide and transversely mag-

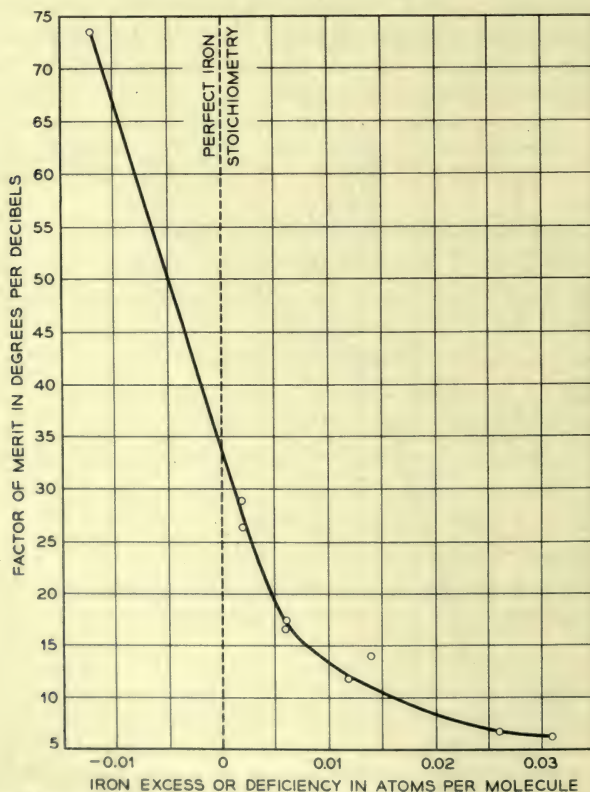


FIG. 19 — Factor of merit versus iron content of NiZn ferrite.

netized. Since these latter effects are quite new two measurements of this type are shown to illustrate the sort of performance obtainable. In the first experiment a slab of ferrite 20 mils thick was placed successively in several positions in a rectangular waveguide, and the phase shift as a function of field was measured for both directions of propagation. When the plate of ferrite is centrally located the phase shift is the same in both directions but when the ferrite is placed half way between the center and the edge of the guide a large difference in phase shift is observed. Finally when the slab of ferrite is located at the edge of the guide there is only a small difference between the positive and negative phase characteristics. This small difference will vanish as the thickness of the ferrite plate goes to zero. In Fig. 21 (a) we show the phase shift versus applied field for three positions of the ferrite slab and in Fig. 21 (b), the differential phase shift at a constant value of applied field is plotted against the position of the ferrite plate. The solid line curve taken at 9,500 mc indicates that maximum differential phase shift is obtained when the ferrite is located approximately 0.100" from the guide wall. Suhl's prediction is that the position at which maximum differential phase shift is observed should be independent of frequency. The dotted curve in Fig. 21 (b) taken at 8,200 mc verifies this part of the prediction in that the maximum again occurs where the ferrite plate is 0.100" from the guide wall even though the point at which $\mathbf{h}$ is circularly polarized has been shifted significantly by the change in frequency.

The second measurement was designed to measure the non-reciprocal absorption which is obtained when the strength of the dc magnetic field is adjusted so that the ferrite is at ferromagnetic resonance. In order to obtain the minimum forward loss and maximum reverse loss it is essential that the ferrite be located precisely at the point where the transverse and longitudinal components of the $\mathbf{h}$ field of the wave are equal, i.e., where the $\mathbf{h}$ field is circularly polarized in a plane perpendicular to the magnetic field. Since this condition exists at only one point in the half-waveguide the ferrite slab must be made very thin. Measurements were made in the 6000–7000 mc band in RG50 waveguide. A thin plate of "Ferramic G" was cut so as to extend from one broad wall to the other. Its length was approximately $1\frac{1}{2}$ " and its thickness was originally 0.050" and was subsequently reduced to 0.025" and finally to 0.009". In the last case the ferrite was so fragile that it was necessary to support it by cementing it to a $\frac{1}{8}$ -inch plate of polystyrene. For convenience the ferrite plate was fastened securely in place at a point calculated to be the point where the $\mathbf{h}$ vector is circularly polarized at the center of the band, and frequency was varied about this center frequency. At each

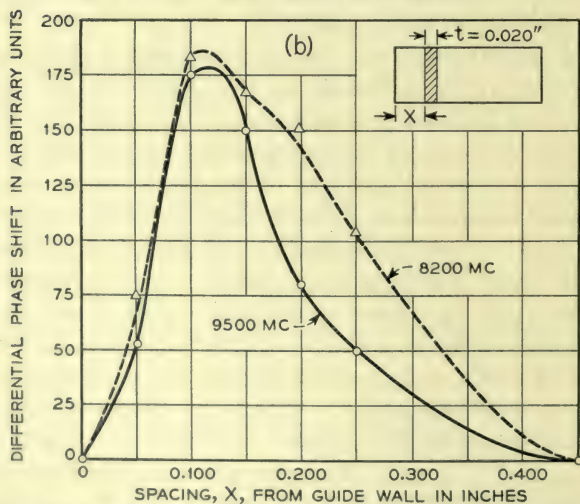
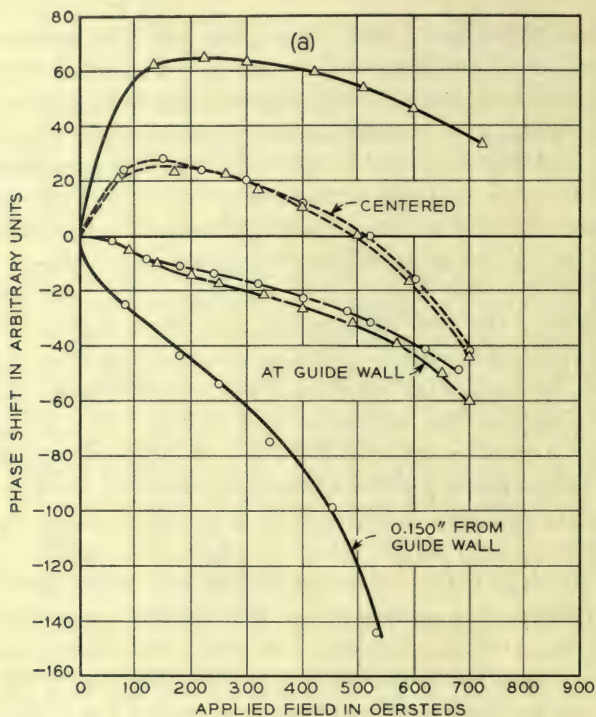


FIG. 20 — Phase shift versus applied field and differential phase shift versus position of the ferrite plate.

frequency the dc magnetic field was adjusted so that absorption was a maximum for the wave for which the absorption is the greater. The field was then reversed and the absorption was again measured. Since reversing the field is entirely equivalent to reversing the direction of propagation we obtain in this way the insertion loss for both directions of propagation. Data taken in this manner are shown in Fig. 22 along with a sketch of the geometry employed. The VSWR was measured in the case of the thinnest plate and was found to be less than 1.05 to 1 over the band. The variations in the dotted curve, however, are probably due to reflections from the ends of the sample.

One must bear in mind that as the frequency is changed, the field required for ferromagnetic resonance is changed so that these data do not give a true index of the bandwidth of the device. This must be measured at a constant value of field. However, if we had used a ferrite

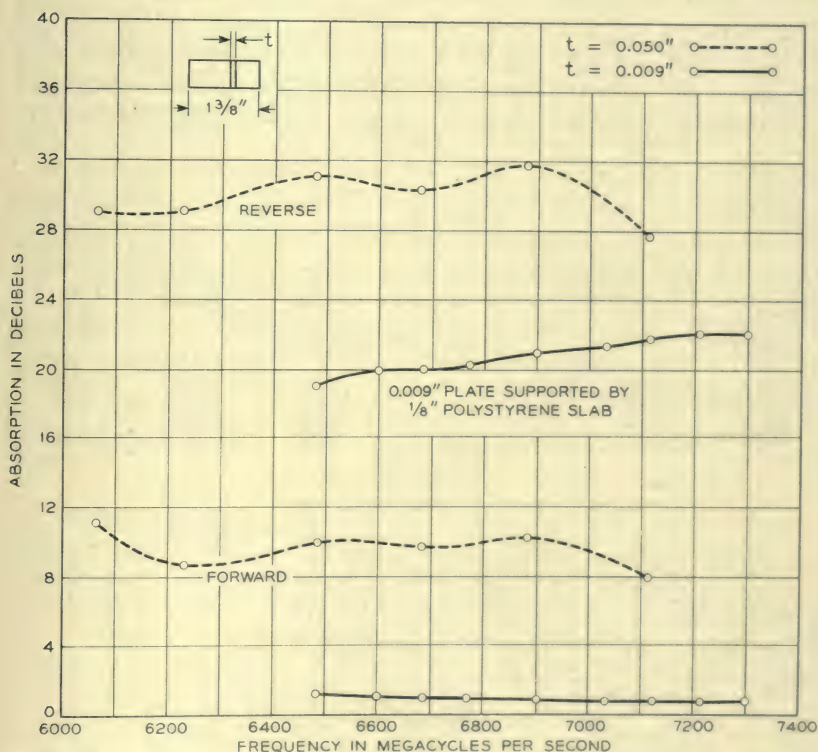


FIG. 21 — Absorption of forward and reverse waves at the optimum value of field as functions of frequency. Two sample thickness are shown to illustrate the effect of sample thickness.

having a loss characteristic such as is shown in Fig. 5, Curve C, in which the loss differential becomes great as soon as the material is saturated, then the bandwidth would be as shown in Fig. 21. Of all the materials which have been measured in the past two years only one sample has shown this effect. At the time such behavior was regarded as the antithesis of the ideal, and no further investigation was made. Now that there is a use for such a material, effort is being directed toward maximizing the effect. Since the effect is thought quite definitely to be due to a very high effective anisotropy field arising either from crystalline anisotropy or demagnetizing effects, the problem of creating the proper material should be quite straightforward.

At this point we should consider the relative merits of the transverse field non-reciprocal devices and those employing the Faraday rotation. We have seen that it is possible to construct a simple isolator using the non-reciprocal absorption in rectangular waveguide. Such an isolator has a minimum forward loss of more than 0.5 db for 20 db reverse loss where an isolator employing the Faraday Rotation can be made with less than 0.1 db forward loss for 30 db return loss. On the other hand the transverse field isolator is much simpler, more compact and easier to

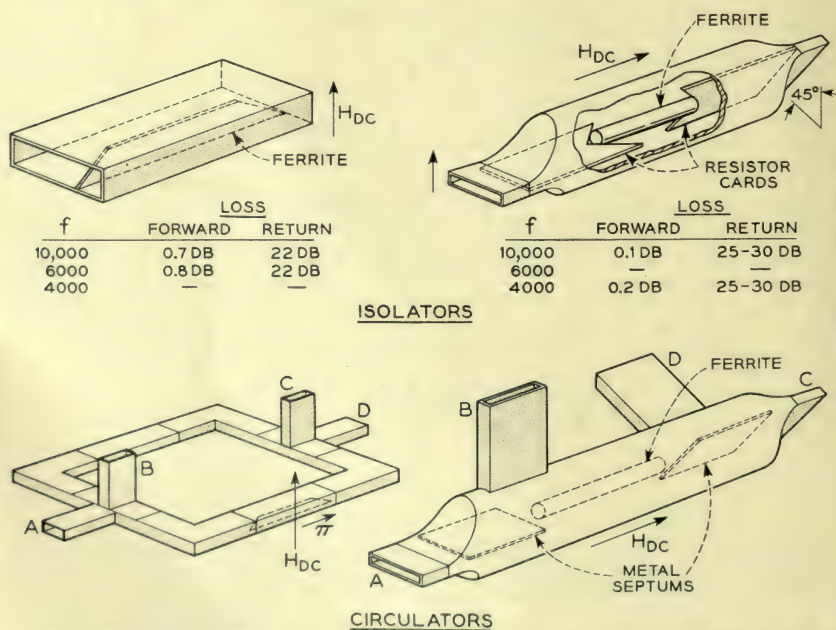


FIG. 22 — Comparison of two basic non-reciprocal elements of the Faraday effect and transverse field effect types.

match impedance-wise. To illustrate the comparison between comparable devices using each principle we have prepared a figure showing the relative structural complexity and performance standards of each. This comparison is shown in Fig. 22. In general the transverse field devices which depend on differential phase shift have about the same insertion loss as those employing the Faraday effect, while the transverse field devices which depend upon differential absorption are somewhat more lossy but simpler to construct than the corresponding Faraday Rotation devices.

SUMMARY

This paper has reviewed the plane wave theory, extended it to discuss waveguide effects, analyzed the various loss mechanisms present in ferrites at microwaves, and discussed numerous measurement techniques and results. It is known that there are original papers in preparation in Bell Telephone Laboratories, and possibly elsewhere, which make this review incomplete at the time of writing. However, it is hoped that the information summarized herein will be of assistance to those who are seeking orientation in this new and rapidly expanding field.

In conclusion the author wishes to thank A. G. Fox, M. T. Weiss, J. P. Schafer, H. Suhl, A. D. Perry and L. R. Walker for permission to discuss herein some of their work and ideas which have not previously been published. The cooperation of F. J. Schnettler and L. G. Van Uitert in providing us with a wide variety of excellent ferrites and in suggesting useful variations in ferrite properties has been of tremendous help in the advancement of the art. The advice given by C. L. Hogan, J. K. Galt and H. Suhl in discussions has proved invaluable and finally the author wishes to thank J. L. Davis for the able assistance he rendered.

Cold Cathode Tubes for Transmission of Audio Frequency Signals

By M. A. TOWNSEND and W. A. DEPP

(Manuscript received August 13, 1953)

Cold cathode gas filled tubes have been extensively applied as electronic switching elements in the telephone system. In general, these applications have been limited to control circuits. The usefulness of these tubes can be further extended by making them capable of carrying voice frequency signals. The transmission properties that are required of the tube for this use are considered. It is shown that troublesome oscillatory noise can be eliminated and that the insertion loss of the tube can be reduced to a low value. Furthermore, by a special design of cathode a stable insertion gain of a few db may be realized. Other requirements on bandwidth, power and distortion are satisfactorily met. Thus, these tubes are potentially useful in coordinate type switches in which voice frequency signals must be rapidly switched.

INTRODUCTION

Cold cathode glow discharge tubes have found increasing usefulness as two-valued switching elements in telephone and other automatic control systems. In many applications the tube functions as a simple control element which either does or does not pass current, depending on the control signals applied. The output signal is in the form of a voltage or a current which can be used to trigger other tubes or operate relays.¹

In other types of circuits the glow discharge tube may be used as a switch in series with a transmission path for audio frequency signals. When used for this purpose, the tube not only must fulfill the switching requirements but also must meet an additional set of requirements which may be realized by controlling the dynamic properties of the discharge. This paper describes some of the characteristics of cold cathode tubes which have been developed for use as switching elements in series with voice frequency telephone transmission circuits.

TRANSMISSION REQUIREMENTS OF AN ELECTRONIC SWITCH

When an electronic switch is used as a substitute for a pair of metallic contacts, a number of requirements must be met in order that voice

frequency currents be faithfully transmitted. These are as follows:

1. *Gain or Loss* — The impedance level of telephone voice frequency circuits is usually of the order of several hundred ohms. The impedance of any device inserted in this circuit should be sufficiently small compared to this value in order to restrict the insertion loss to less than 1 or 2 db. Of course, the impedance levels of the circuits may be raised to higher values. Aside from the extra cost of transformers, the problems of noise pickup and crosstalk become bothersome when this is done. Since no amplification of the signal from the telephone set is normally used in local transmission circuits, any value of loss is highly undesirable. In fact, since the use of the electronic switch may require the introduction of other circuit elements which introduce loss, it would be desirable that the electronic switch provide a small amount of gain.

The above discussion has considered the gain or loss of the switch in its "closed" condition. Between two busy circuits, there may exist paths consisting of one or more "open" switches. Each of these paths may contribute crosstalk into the circuits. Therefore, impedance required of an individual switch must be high. For a large coordinate switch, there will be a very large number of these undesired paths. This requires that the open impedance of the individual switch be of the order of hundreds of megohms.

2. *Bandwidth* — It is desirable that the electronic switch transmit faithfully frequencies of 300 to 3,500 cycles per second.

3. *Power Output and Distortion* — Since the impedance of the electronic switch may vary with the current passing through it, distortion may be introduced when the current swings are large. On the other hand, without the proper current swing, insufficient power will be delivered to the load impedance. Telephone circuits need to handle powers of the order of a few milliwatts with a harmonic distortion less than a few per cent.

4. *Noise* — The noise introduced into a telephone switching system by an electronic switch should not be noticeable to a subscriber. This means it should be below 10^{-8} or 10^{-9} watts.

5. *Stability* — The properties that have been considered above must be highly stable with time. In central office use, such devices might be used for periods of ten to forty years.

STATIC CHARACTERISTICS

A common form of glow discharge tube comprises a pair of metal electrodes in a glass envelope which is carefully evacuated and filled with

a chemically inert gas to a pressure ranging from 1 to 100 mm of mercury. The negative electrode, called the cathode, is given a special processing which permits it to emit electrons readily when bombarded with positive ions of the gas. The positive electrode, the anode, serves to collect electrons emitted by the cathode as well as those produced in the gas by ionization.

A typical voltage-current characteristic is shown in Fig. 1. For discussion purposes we may divide the curve into several current ranges. In current range I a small residual current flows even at low voltages because of ionization resulting from cosmic rays or radio-active material placed in the tube. The two curves in range I are for different residual currents. At higher voltages this residual current is amplified as a result of additional ions and electrons formed in the gas but it is still extremely small. If the tube voltage is increased still more, the current increases very rapidly until in current range II, a self-sustaining discharge is established. Each electron released from the cathode gains enough energy on the way to the anode so that it produces a large number of positive ions, excited atoms, and photons in the gas. When these particles, arriving back at the cathode, on the average, release another electron, the

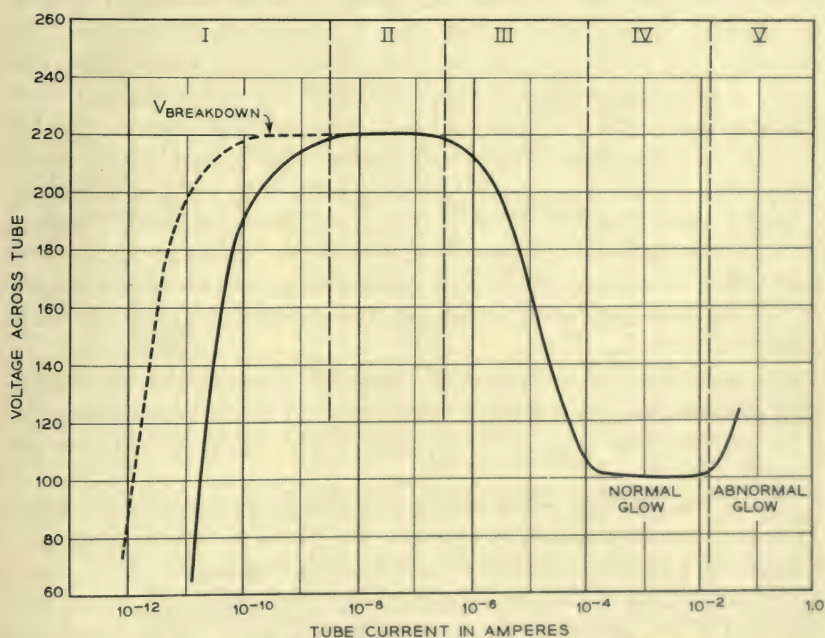


Fig. 1 — Volt-ampere characteristic of cold cathode diode.

self-sustaining condition has been achieved. The value of voltage at which the discharge becomes self-sustaining is usually referred to as the breakdown voltage. A detailed discussion of current ranges I and II has been presented by Druyvesteyn and Penning.³

Because the tube is a good insulator in current range I at low voltage and suddenly becomes a good conductor in range II, we often think of a gas tube as a voltage controlled device and use it as such in switching circuits. Actually, however, in current range II the tube can be controlled only by regulating the current. For this reason in the remainder of this paper, the current is considered as the independent variable.

In current range III the voltage falls rapidly with increasing current. A space charge of positive ions begins to develop close to the cathode. By the beginning of current range IV most of the voltage drop in the tube appears across this space charge layer. This region of nearly constant voltage with increasing current is called the normal glow discharge. The space charge layer immediately in front of the cathode is commonly called the cathode dark space. Electrons emitted from the cathode as a result of positive ion bombardment and other processes are accelerated through the high field of this cathode dark space and produce an adjoining layer of intense ionization and excitation called the negative glow. In tubes of the type considered here, this negative glow is the most luminous part of the tube. Beyond the negative glow toward the anode is the so-called Faraday dark space in which no new ionization or excitations are produced. Electrons from the negative glow can readily flow through this region to the anode because their space charge is almost completely cancelled by positive ions diffusing from the negative glow.

Over current range IV the cathode current density is nearly constant. This means that the cross-sectional area of the discharge increases in proportion to the current. This is evidenced by the familiar spreading of the negative glow with increasing current until it covers the entire cathode area.

In current range V the cathode is completely covered with the negative glow and the current density must increase in direct proportion to the total current. This range of currents is called the abnormal glow discharge.

The current-voltage characteristic of a cold cathode for current ranges IV and V may be modified by changing the geometry from a single plane cathode to a hollow cathode.³ A hollow cathode is one in which there is an overlapping of the regions of cathode fall and negative glow from two portions of the cathode. This overlapping can occur on the inside of a cylindrical or spherical surface or between two plane cathodes more or

less parallel to each other and closely spaced. The optimum dimensions are a function of the density and kind of filling gas used. Electrons and ions generated in the cathode fall and negative glow regions of one cathode can be more efficiently used in producing new electrons and ions if they can enter another cathode fall region instead of diffusing outward into the Faraday dark space. This effect is particularly noticeable in the abnormal glow range of currents.

This is illustrated by the curves of Fig. 2 which were taken on a tube containing two parallel plane cathodes with an adjustable spacing between them. Because Fig. 2 is plotted to a linear current scale, current ranges I and II of Fig. 1 are compressed to the left-hand axis. The upper curve is obtained when the cathodes are far apart so that there is no interaction. The lower curve is obtained when the spacing is close enough to give a hollow cathode effect.

The sustaining voltage of a normal glow discharge is dependent upon the anode to cathode distance. To investigate this we can arrange a parallel plane cathode-anode structure with the anode attached to a piece

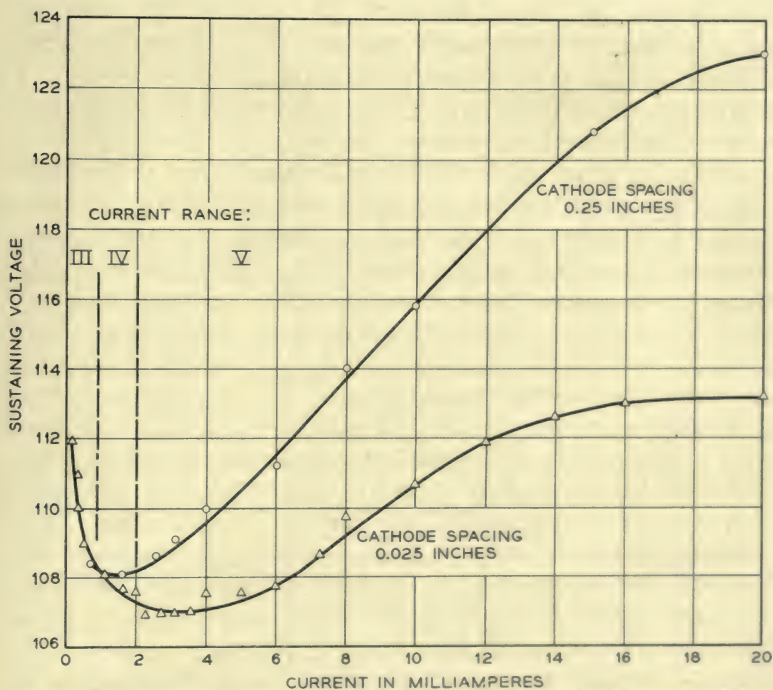


Fig. 2 — Volt-ampere characteristic of parallel plane cathodes at two different spacing.

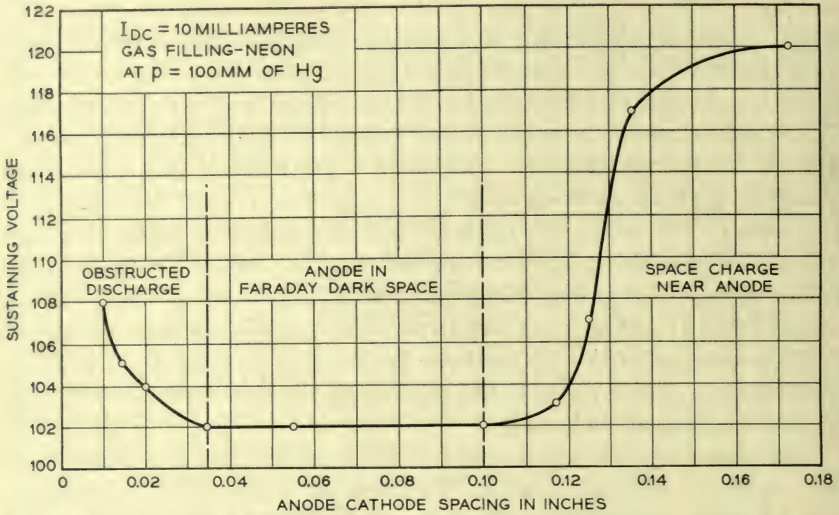


Fig. 3 — Sustaining voltage as a function of anode-cathode spacing.

of magnetic material which can be moved by a magnet external to the tube. Fig. 3 shows a typical set of data obtained in this manner with an operating current in the normal glow range of current. At very close spacings of the order of the negative glow distance, the voltage is higher than normal because some of the electrons released from the cathode are able to strike the anode before dissipating their energy in producing ions and excited particles in the gas. This is referred to as an "obstructed discharge".³ The voltage must, therefore, be higher because electrons which do produce excitation or ionization must be given extra energy in order to maintain the current. As the anode is moved away from the cathode through the Faraday dark space, the tube sustaining voltage stays at a fairly constant minimum value. This is possible because at close spacings more than a sufficient number of electrons and ions are present in the Faraday dark space to carry the current. As the anode is moved away the Faraday dark space is lengthened. Ions and electrons needed to carry the current diffuse from the cathode region. At sufficiently large distances, however, the ionization density decreases so that not enough ions are present to cancel space charge near the anode. A space charge sheath builds up in the anode region and the sustaining voltage rises with increasing distance.

When the voltage has increased by about 10 to 18 volts depending upon the gas filling, it begins to level off with increase in distance. At about this point, an anode glow may appear in front of the anode. This is due

to the fact that some of the electrons have gained sufficient energy for excitation. A slight further increase in anode-cathode distance usually results in the anode glow changing from a uniform layer to a "ball" of glow. When this occurs, oscillations of several volts amplitude appear across the tube terminals. These oscillations result from a sequence of events which is initiated when the electrons gain enough energy in passing through the anode fall region so that they may ionize. A small number of ions generated in this region will, because of their relatively low velocity, enable a large electron current to flow without developing space charge. This then will reduce the voltage appearing across the anode fall and greatly reduce the number of ionizations taking place. As soon as the recently produced ions leave this region the voltage drop across this region increases causing the ionization to build up again. This alternate building up and decaying of ion density results in the observed oscillation which is ordinarily in the frequency range from 0.5 to 20 kilocycles per second.

This oscillation usually cannot be prevented by external circuit means. However, by proper choice of anode-cathode spacing, type and density of the gas filling, and to lesser extent the geometry of the anode, a tube can be made which is free from anode oscillations. The main restriction that this puts on tube design is the limitation of breakdown voltage.

IMPEDANCE

From the previous discussion of transmission requirements it is clear that one of the most important properties of a gas tube is the impedance presented to small ac signal currents superimposed on the steady dc operating current. At low frequencies, these signals cause the voltage across the tube to vary in accordance with the static characteristic. The impedance of the tube to these signals is almost entirely resistive and is equal to the slope of the static characteristic. At higher frequencies, however, there is a lag in the adjustment of the voltage across the discharge to the changes in current. Hence, at these frequencies, the impedance of the tube may have both resistive and reactive components. This is illustrated in Fig. 4.

The small superimposed signals result in current-voltage loci which are ellipses. In current range III at 200 cps, the position of the ellipse corresponds to a negative resistance in series with an inductance. The negative resistance changes rapidly with frequency and as shown at 2,000 cps, it may be positive. Because of the rapid variation of impedance, both with frequency and current, this range of currents is not generally useful for dependable transmission of voice frequency signals.

In the higher current range IV and V the impedance can be represented by a positive resistance in series with an inductive reactance. The resistance increases slowly with frequency as indicated by the elliptical current-voltage loci at 200 and 2,000 cps.

The impedance of a cold cathode tube also varies with anode-to-cathode spacing. This may be studied by means of the same movable-anode parallel-plane tube used to obtain the data of Fig. 3. We again operate the tube in the current range of the normal glow (IV) and, for illustrative purposes, choose a measuring frequency of 1,000 cycles per second. The results are shown in Fig. 5. The resistive and reactive components of tube impedance are independent of anode spacing throughout the Faraday dark space and well into the obstructed discharge region. At large distances where the electron space charge sheath begins to build up in front of the anode, the impedance increases rapidly with distance.

Some useful conclusions about the design of transmission tubes can be drawn from the data of Fig. 5. We can consider the total tube impedance as being made up of the sum of the impedances introduced by the various regions of the discharge. Since large variations in the length of the Faraday dark space do not affect tube impedance, it is concluded that this region has negligible impedance. This means that, so long as the anode-to-cathode distance is short enough to avoid a space charge sheath at the anode, we can concentrate our attention on the cathode fall region. The following detailed discussions of impedance are consequently restricted to the cathode portions of the glow discharge.

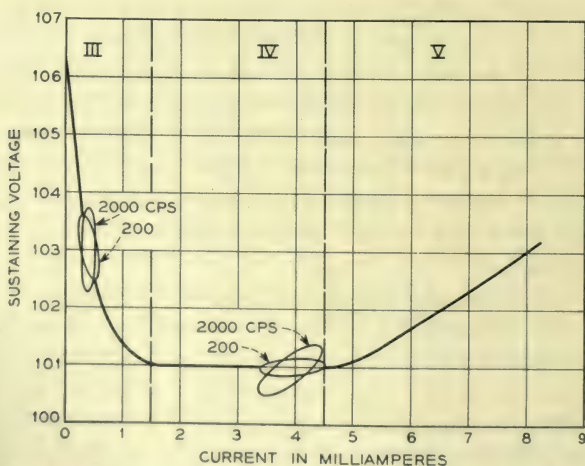


Fig. 4—Static and dynamic volt-ampere characteristics.

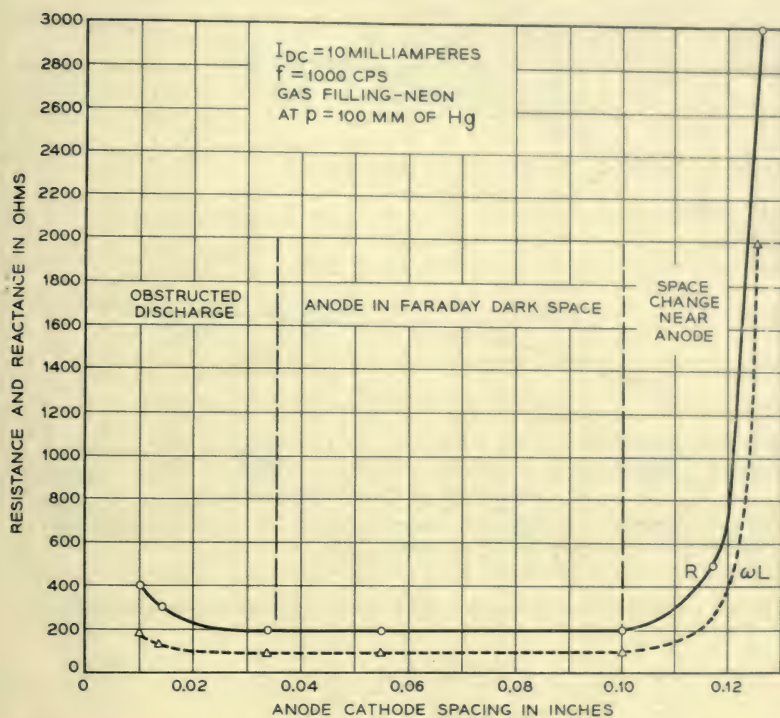


Fig. 5 — Resistive and inductive components of impedance as a function of anode-cathode spacing.

IMPEDANCE OF PLANE CATHODE TUBES

Let us consider first the impedance of a tube with a plane cathode or of a cathode with a radius of curvature large compared to the combined thicknesses of the cathode dark space and the negative glow. An example is the Western Electric 313-C which has a gas filling of 99 per cent neon, 1 per cent argon and a cathode surface of barium and strontium oxides. At a fixed steady current of 25 ma flowing through the starter gap, it is found that the resistive component of the impedance varies with frequency as shown in Fig. 6. In the middle of the voice band it has a value of about 200 ohms. There is also an inductive reactance of about 65 ohms at this frequency. At a fixed frequency, it is found that the resistive component of this type of tube decreases with current approximately inversely as the square root of the current. This is shown in Fig. 7. The

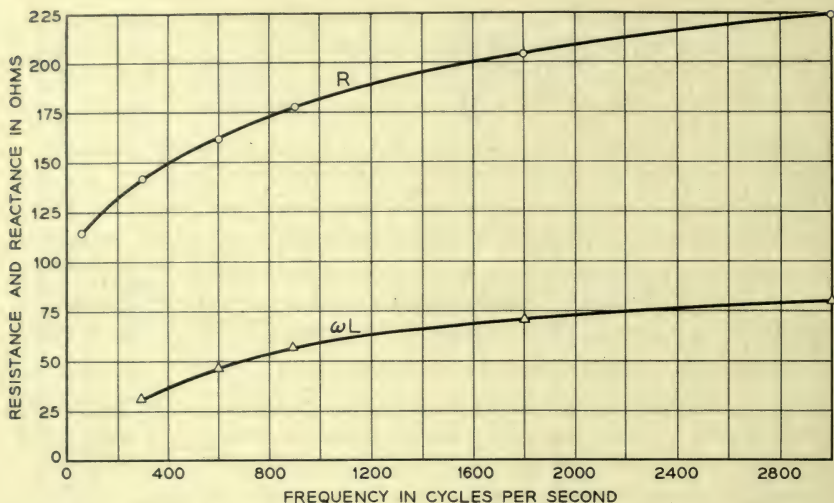


Fig. 6 — Resistive and inductive components of 313-C starter gap impedance as a function of frequency. $I_{dc} = 25$ mm.

inductive reactance is related to the current approximately as follows:

$$\omega L = B + \frac{C}{I_{dc}}. \quad (1)$$

B and C are functions of frequency.

If we wish to use tubes of this type in central office switching circuits which pass voice frequency currents, we find that even one tube connected between a 600-ohm source and a 600-ohm load gives a prohibitively high insertion loss. The impedance level of the voice frequency circuit may, of course, be raised up to 4,000 or 5,000 ohms. But practical switching circuits might require as many as four tubes in series. Hence, we see that the impedance of tubes of this type severely limits their usefulness.

From Fig. 7 we might argue that to get low resistive components we might continue to increase the direct current flowing through the tube. However, in this range of currents, the life of the tube varies approximately as $1/I_{dc}^3$ while the resistive component varies as $1/I_{dc}^{1/2}$.

Over the range of currents plotted in Fig. 6, the cathode is fully covered with glow. As noted above, the resistive component is

$$R = \frac{A}{I_{dc}^{1/2}}. \quad (2)$$

If instead of passing a given direct current through a single tube we pass the same current through a parallel combination of n tubes, the resistive component of the combination would be

$$R \text{ comb.} = \frac{1}{n} \frac{A}{\left(\frac{I_{dc}}{n}\right)^{1/2}}, \quad (3)$$

$$= \frac{1}{n^{1/2}} R.$$

The resistive component of the combination is $\frac{1}{n^{1/2}}$ times that of a single tube passing the same total current. This assumes, of course, that the cathodes of the tubes are covered with glow. Obviously this is not a practical method of attaining a low impedance because of the instability of paralleled individual tubes. It does suggest, however, that by increasing the cathode area of a single tube until the glow just covers the full area, a lower impedance is obtained. This has been done experimentally.

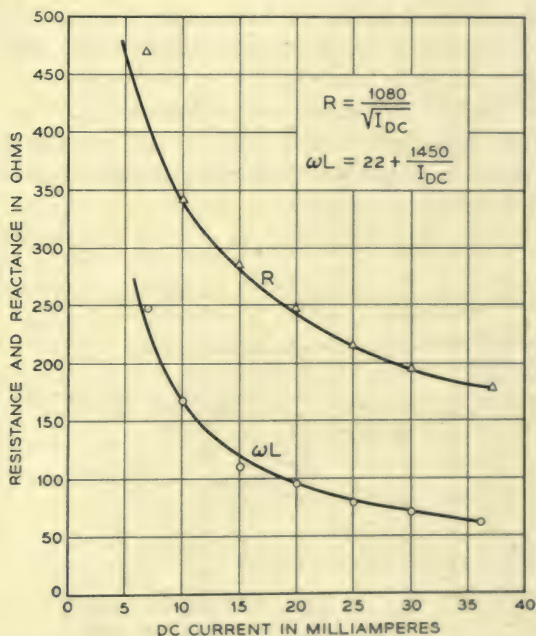


Fig. 7 — Resistive and inductive components of 313-C starter gap impedance as a function of direct current.

Fig. 8 shows the data for a tube of this type which has a cathode area about six times that of the 313-C. This change in cathode area along with the necessary change in anode geometry led to an impedance reduction of about five times. This type of design, of course, gives a large increase in life of the tube but it is undesirable from the standpoint of tube size.

As pointed out above, low impedance can be attained by an increase of the tube current and within limits by an increase of the cathode area. A third parameter which may be varied is the density of the gas filling. At a constant current and with a given cathode area, the resistive component of the impedance decreases with increase in density or the pressure, p , at a fixed temperature approximately in accordance with the relation

$$R \propto \frac{1}{p^2}. \quad (4)$$

The curve in Fig. 9 was obtained over a limited range of argon fillings with a barium strontium oxide cathode. Since the current density increases approximately in proportion to p^2 , the effective cathode area will tend to be reduced unless the total current is kept high enough to cover the cathode fully. Therefore, with a fixed total current there is a limit to either increase in pressure or in cathode area.

In further search for low impedance, a number of different gas fillings have been tried. They have included all the rare gases as well as several mixtures of them. No significant advantages were obtained by the use of other than the more common neon or argon gas fillings.

We therefore see that the ability to control impedance properties of

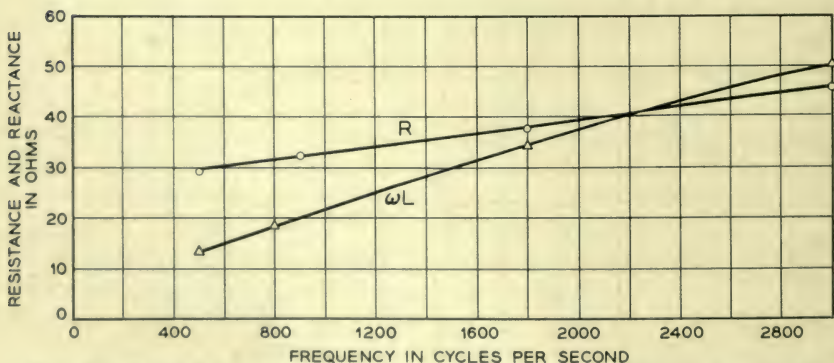


Fig. 8 — Resistive and inductive components of impedance for a large area cathode tube as a function of frequency. $I_{dc} = 25$ mm.

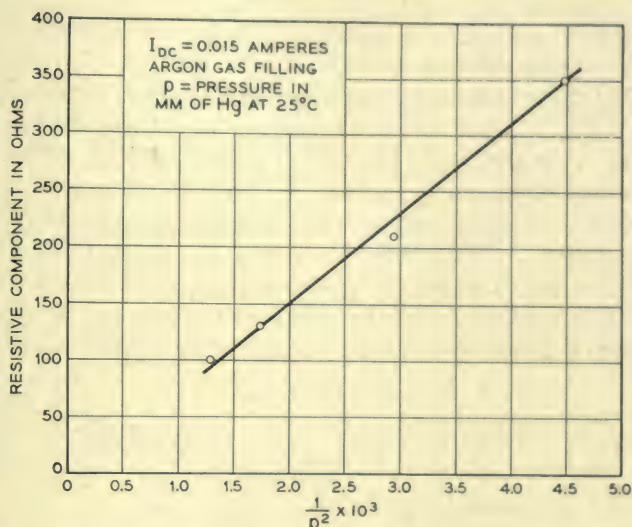


Fig. 9 — Resistive component of impedance as a function of density (pressure at 25°C) of gas filling.

the plane cathode through changes in tube current, cathode area, density and type of gas filling is limited. A reduction by a factor of five or ten times over the values present in commercial tubes is certainly possible. But the possibility of obtaining values of less than +10 ohms or negative values seemed remote. Consequently, development effort was concentrated on the hollow cathode tube described below.

HOLLOW CATHODE TUBES

The static voltage-current characteristic of a hollow cathode was shown in Fig. 2 to be below that of a plane cathode of the same area when operating at currents in the abnormal glow region. It has been found that by proper choice of the cathode dimensions and the kind and density of the filling gas, desirable transmission characteristics can be achieved. The following discussion illustrates the manner in which some of the variables are interrelated.

We will consider a "U" shaped cathode which has been formed by folding a piece of molybdenum sheet in the form illustrated in Fig. 10. The choice of dimensions of the hollow portion will be discussed later. The anode is a cylindrical rod placed in front of the cathode. The anode-to-cathode spacing is selected so that the anode is always within the Faraday dark space and hence does not influence the impedance appreciably. It is assumed that the structure of Fig. 10 is sealed in a bulb

which has been carefully evacuated and filled to a fraction of an atmosphere with a rare gas such as neon.

A typical static volt-ampere characteristic of the structure of Fig. 10 is shown in Fig. 11. This curve was taken with a cathode having a hollow portion $\frac{1}{8}$ " long and $\frac{1}{16}$ " deep with a cathode gap of 0.023". A neon filling pressure of 58 mm of mercury was used. It can be seen that there is the usual low-current negative slope associated with the transition from breakdown (current range III of Fig. 1). A new characteristic of interest is the second region of negative slope in the abnormal glow range of currents. It has been found that this second region can be made stable with time in a given tube and reproducible from tube to tube. It is also

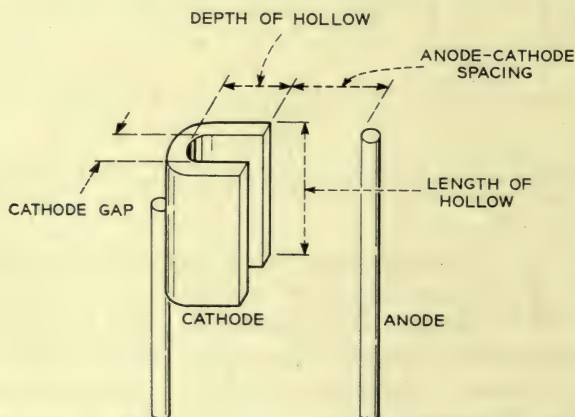


Fig. 10 — Electrode geometry of a hollow cathode tube.

found that the tube impedance has a negative resistance component over the voice frequency range. Thus, this second region of negative resistance offers attractive possibilities as a transmission element.

The impedance of this tube at 300 and 3,000 cps is shown in Fig. 12 for the same range of operating current as Fig. 11. The optimum current for negative resistance and the value of negative resistance are functions of the cathode gap, but so long as the other cathode dimensions are constant the optimum current is relatively independent of the density of the filling gas.

A useful way of studying the interrelation of cathode gap and filling pressure is shown by Figs. 13 and 14. For these data the length and depth of the hollow portion were kept constant at the values of $\frac{1}{8}$ " and $\frac{1}{16}$ " respectively. Fig. 13 shows the resistive component of impedance as a function of frequency for different filling pressures with a fixed cathode gap.

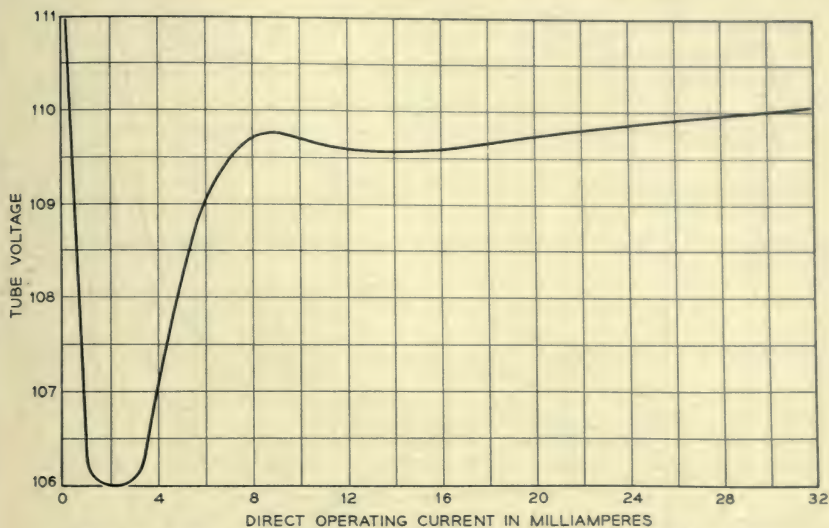


Fig. 11 — Static volt-ampere characteristic of hollow cathode tube.

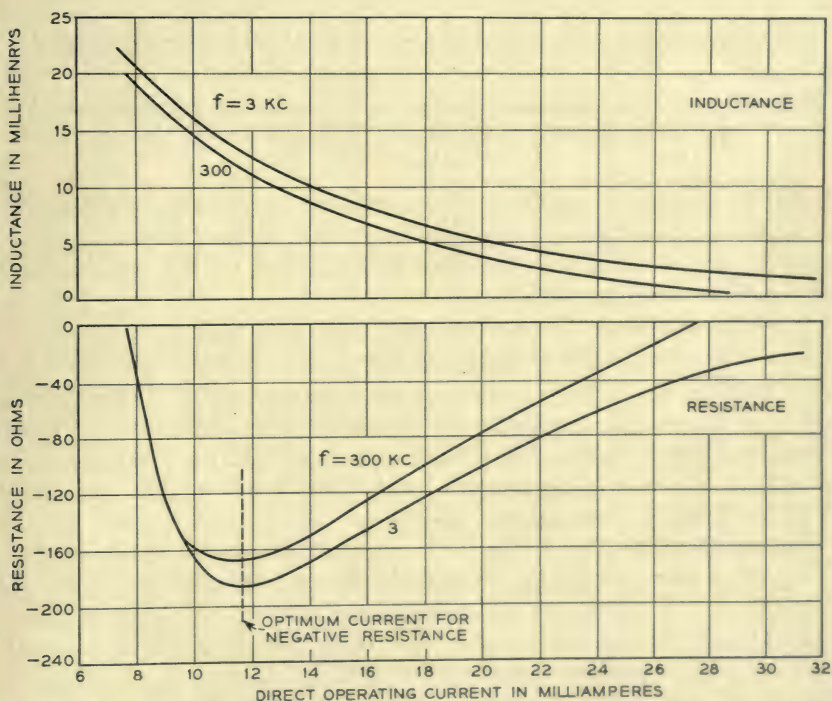


Fig. 12 — Resistive and inductive components of impedance of a hollow cathode tube.

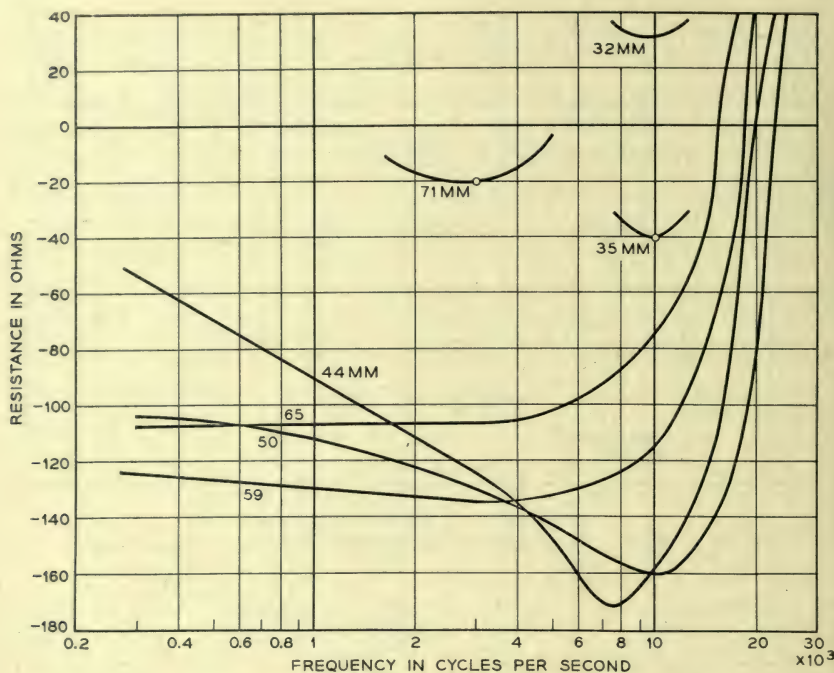


Fig. 13 — Resistive component of impedance as a function of frequency for different neon filling pressures. Cathode gap = 0.024 in.

Fig. 14 shows the variation of the resistive component of tube impedance as a function of frequency for several cathode gaps and at a fixed density of filling gas. For both Figs. 13 and 14, the current was adjusted for optimum negative resistance.

It can be seen from Fig. 13 that the choice of a neon filling gas at a pressure near 60 mm and from Fig. 14 that a cathode gap near 0.024 inch could be expected to yield a negative resistive component of impedance which is reasonably insensitive to filling pressure and which is also constant in value over the voice frequency range. This justifies the choice of cathode gap and filling pressure used in the tube on which the data of Figs. 11 and 12 were taken.

One other parameter of interest is the limit of anode-to-cathode distance. This too is a function of cathode gap and gas density. A typical curve taken for the same cathode geometry and gas filling as used for Figs. 11 and 12 is shown in Fig. 15. It is seen that the negative resistive component is essentially independent of anode distance for a distance of approximately 0.050 inch. This means that the breakdown voltage of

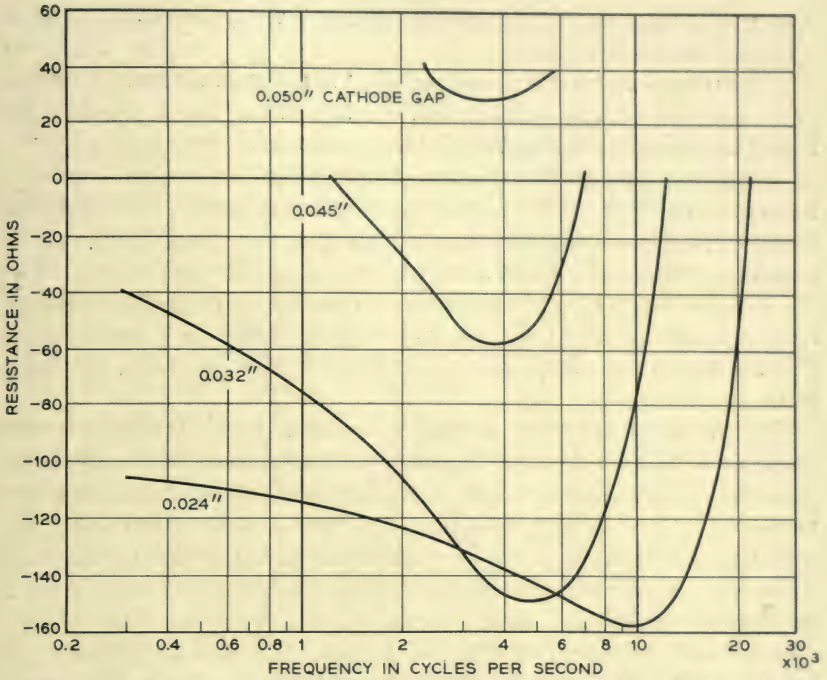


Fig. 14 — Resistive component of impedance as a function of frequency for different values of cathode gap. Neon filling pressure = 50 mm of Hg.

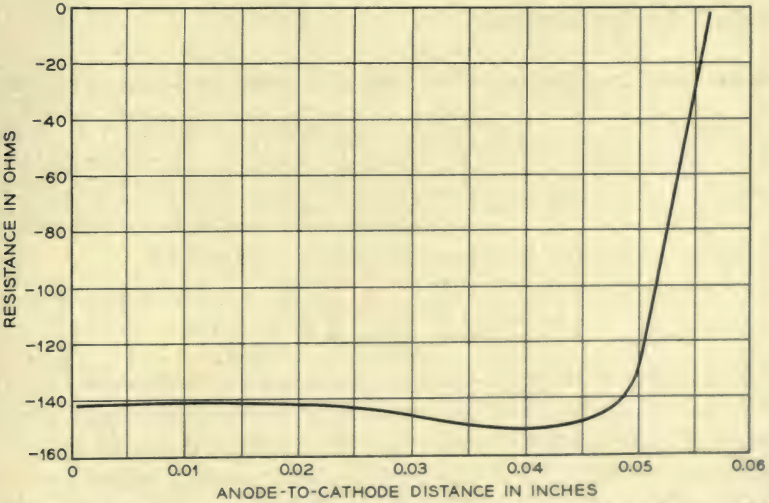


Fig. 15 — Resistive component of impedance as a function of anode-to-cathode distance.

a switching tube can be adjusted by changing this spacing so long as the upper limit is not exceeded.

Although the above discussion is limited to only one adjustable cathode dimension, the cathode gap, other variations are of course possible. The length and depth of the hollow should be at least a few times the width of the cathode gap in order that an efficient hollow be formed. Increasing the hollow length or depth requires larger currents and, if carried too far, the glow may not completely fill the hollow at the optimum current for negative resistance. This is undesirable because the unused portion of the cathode may change its properties with time and produce unstable characteristics. The entire geometry may be scaled to a larger size if the density of the filling gas is reduced by approximately the same factor.

The choice of cathode material is restricted by the high current densities of hollow cathodes. Coatings of alkaline earth oxides or similar materials have too short a life. Pure molybdenum has found to give a satisfactorily low sustaining voltage together with long life and stable operating characteristics. Life tests have shown that tubes can be made which will operate satisfactorily for the equivalent of 20 to 40 years in central office service.

It is seen from the above that by changing the cathode geometry and density of the filling gas a variety of impedance properties can be obtained. The final choice must be determined by the overall transmission requirements. As an example, the transmission performance of a typical tube will now be discussed.

TRANSMISSION PERFORMANCE OF A TYPICAL NEGATIVE RESISTANCE DIODE

The circuit of Fig. 16(a) shows a cold cathode switching tube in series with a transmission path. The voltage across the load resistor R_L under conditions of Fig. 16(a), divided by the voltage across R_L with no tube in the circuit, Fig. 16(b), is one measure of the transmission performance. This ratio, called the insertion voltage gain, is given by

$$\text{I.V.G} = \frac{R_s + R_L}{R_s + R_L + R_t + j\omega L_t} \quad (5)$$

The derivation of this expression assumes that the transformers are ideal and that the reactance of the condenser C is negligible. Maximum gain occurs at low frequency where the reactive component of tube impedance, $j\omega L_t$, can be neglected. If the resistive component of tube impedance, R_t , is negative, the gain will be greater than unity. The gain approaches an infinite value as the unstable condition is approached where the

negative resistance of the tube equals the sum of the source and load resistance. Large values of gain are not practical because this imposes undesirable restrictions on the constancy of the circuit and tube impedances.

Additional restrictions on gain arise from bandwidth and distortion considerations. If it is assumed that both R_t and L_t are independent of frequency over the voice band, it is possible to use the above equation for an approximate calculation of bandwidth. The half-power point occurs at the frequency f_c at which the reactive term equals the sum of the other three terms in the denominator, or

$$f_c = \frac{R_s + R_L + R_t}{2\pi L_t} \quad (6)$$

Since the gain does not fall off at low frequency, the upper cut-off frequency f_c is a measure of the bandwidth. Substituting R_s and R_L from equation (5) into equation (6) gives

$$f_c \text{ (1 - Low frequency I.V.G.)} = \frac{R_t}{2\pi L_t} \quad (7)$$

Thus for a given tube an increase in gain is accompanied by a decrease in bandwidth.

As shown in Fig. 12 the impedance of a negative resistance tube is dependent on the current passing through it. This will cause some distortion as the signal current swings above and below the average direct current value. The distortion is small so long as the non-linear tube resistance is small compared to the total circuit impedance.

It can be seen from the above discussion that for a given tube the gain, bandwidth, and distortion are all dependent on the source and load impedances. Fig. 17 shows the experimental performance of a typical tube for the special case where source and load impedances were equal. Insertion gain has been converted to power gain in db. The distortion

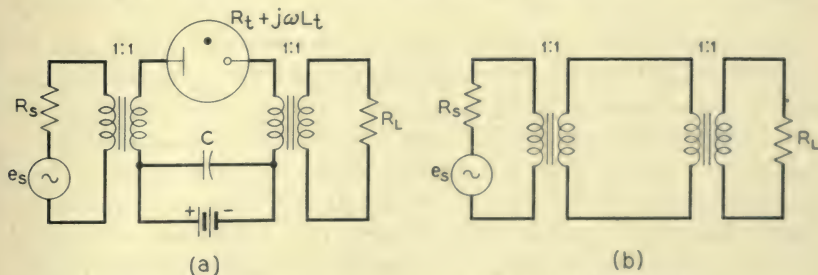


Fig. 16 — Transmission circuits with and without tube.

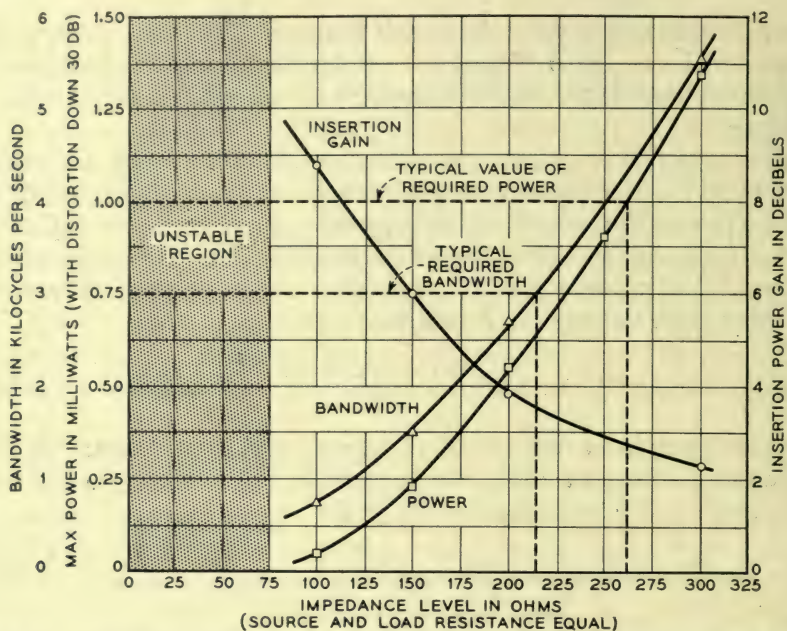


Fig. 17 — Performance curves for experimental cold cathode negative resistance diode.

has been related to the power handling capacity by measuring at each value of source and load impedance, the maximum power output at which the total harmonic distortion is 30 db below the signal.

The noise introduced by a tube affects its usefulness as a transmission element. By designing the tube so that the anode is in the Faraday dark space, anode oscillations are avoided. The remaining noise is at a low level, typical values being 10 decibels above the noise reference level of 10^{-12} watts.

CONCLUSIONS

Cold cathode glow discharge tubes can be made with stable and reproducible impedance characteristics. By proper choice of anode-cathode spacing and pressure of filling gas, it is possible to eliminate oscillation noise associated with the anode region. By properly choosing the density of filling gas and the area of a plane cathode, it is possible to obtain a low positive resistance component of impedance. By proper correlation of cathode geometry and filling gas density, hollow cathodes can be used to obtain a negative resistance component of impedance. Bandwidth, power, distortion, and noise requirements of voice frequency transmis-

sion circuits can be satisfied without sacrificing switching characteristics. Cold cathode glow discharge tubes are, therefore, a potentially useful electronic switch for use in series with voice frequency transmission circuits.

ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of F. T. Andrews in analyzing and measuring the amplifier performance of negative resistance gas tube diodes.

REFERENCES

1. Ingram, S. B., Elec. Eng. (A.I.E.E. Transactions), **58**, pp. 342-346, 1939.
2. Depp, W. A., and W. H. T. Holden, Elec. Mfg., **44**, pp. 92-97, 1949.
3. Druyvesteyn, M. J. and F. M. Penning, Revs. Mod. Phys., **12**, pp. 87-174, 1940.

Balanced Polar Mercury Contact Relay

By J. T. L. BROWN and C. E. POLLARD

(Manuscript received June 19, 1953)

A new type of relay making use of solid contacts maintained continuously wet with mercury has been developed. It has a symmetrical polar structure, resulting in improved sensitivity and speed compared with the existing neutral structure with similar contacts. It is also particularly well adapted to switching of high frequency circuits. Two magnets are used for polarization, and the relay is adjusted after assembly to desired values of sensitivity for operation in both forward and reverse directions by selective adjustment of the magnet strengths.

INTRODUCTION

In a previous paper¹ a mercury contact relay is described in which the contact elements are maintained wet with mercury through a capillary path to mercury reservoir below the contacts. The present paper describes a new design making use of this same mercury contact principle, but with a symmetrical polar structure which gives improvement in sensitivity and speed capabilities over the previous neutral structure.

VERTICAL RELAY

Fig. 1 shows one design of the relay, adapted for general use in a vertical position. It provides a single pole double throw magnetic switch in a sealed glass tube, enclosed along with an operating coil and polarizing magnets in a steel can with a medium octal plug base similar to the previous type relay (275 type).

Fig. 2 shows the glass enclosed magnetic switch element. The armature is a tapered reed welded to a tubular stem which is sealed in the glass at the lower end. Mercury, and gas under pressure are introduced through this tube, which is then welded closed. The magnetic working gaps are formed between the armature and fixed pole-pieces which are sealed in

¹ J. T. L. Brown and C. E. Pollard, Mercury Contact Relays, Elec. Eng., **66**, Nov., 1947.

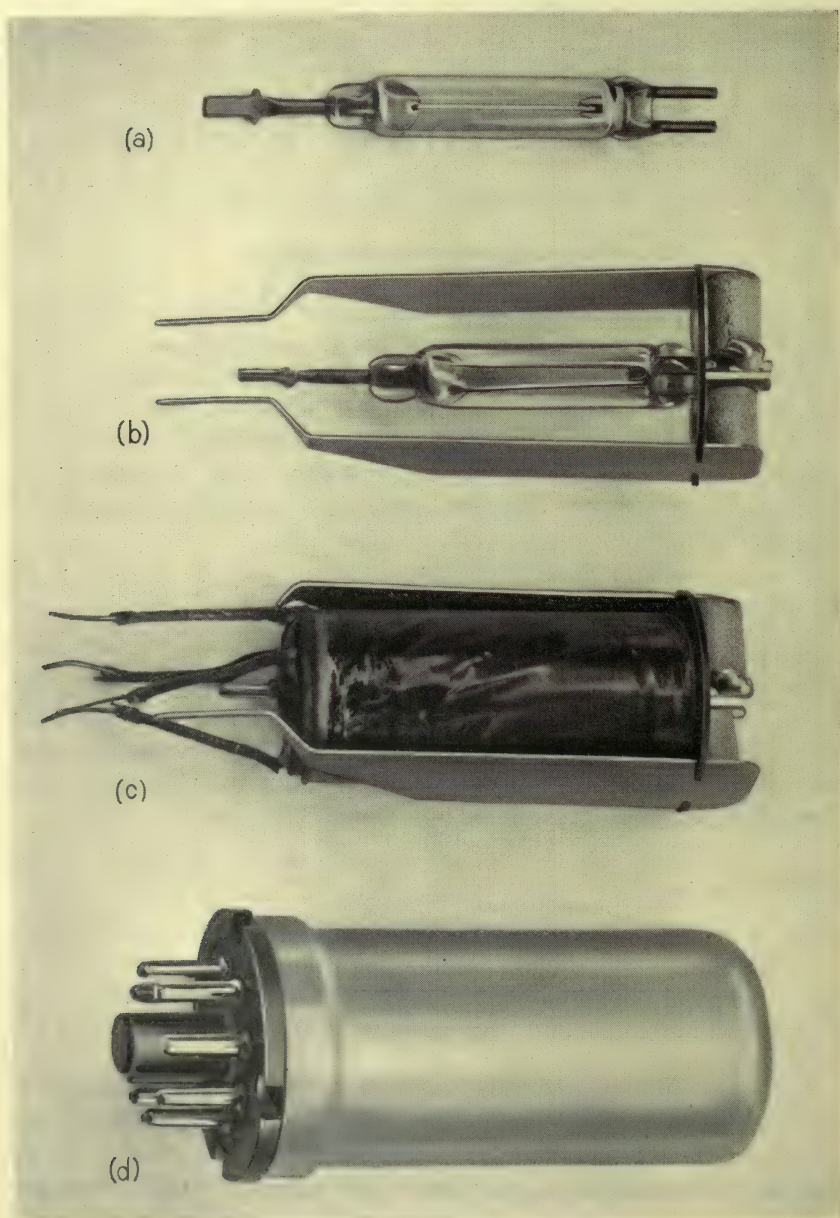


Fig. 1 — (a) Switch element. (b) Magnets, spool, and side plates added. (c) Coil added. (d) The complete relay.

the glass at the upper end. The fixed contacts are made from small platinum alloy balls which are welded to the pole-pieces. The armature strikes the fixed contacts at a point which is close to a node for one of its principal modes of vibration. This limits the bounce on impact to an amplitude which does not open the mercury bridge which is formed.

Mercury in a pool at the bottom of the switch is fed to the contact area through grooves rolled in the armature surface. Except for the platinum alloy contacts, the surfaces of the pole-pieces inside the switch have an oxide coating which prevents them from being wet by the mercury. This limits the mercury bridge which is formed between the armature and pole-piece to the area of the platinum alloy contact. A ceramic detail is inserted between the pole-pieces at the top of the switch. The ceramic is specially resistant to wetting by the mercury and thus prevents mercury from collecting between the pole-pieces.

The polarizing magnets are soldered to the pole-piece terminals outside the switch. Permalloy plates are soldered to the outer poles of the magnets and extend down on the outside of the coil, forming a return path to the lower end of the armature. The coupling at the lower end is made relatively loose in order to limit magnetic soak effects. The steel can provides a magnetic shield.

STATIC MERCURY CONFIGURATION

The static configuration of the mercury surfaces in the switch depends upon the shape and contact angle to mercury of the solid surfaces with which it comes in contact, and the curvature of the free mercury surfaces. This curvature depends on the surface tension of the mercury surface T and the pressure difference Δp between the inside and outside of the mercury at the point under consideration. That is,

$$\Delta p = T \left(\frac{1}{R_1} + \frac{1}{R_2} \right), \quad (1)$$

where R_1 and R_2 are principal radii of curvature of the surface (radii taken in planes cutting the surface at right angles to each other).

In this switch the pressure difference is a function of the height of the point under consideration above the surface of the reservoir:

$$\Delta p = \rho gh. \quad (2)$$

Substituting the values

$$\begin{aligned} \rho &= 13.6 \text{ g/cc for mercury,} \\ g &= 980 \text{ cm/sec}^2, \text{ and} \\ T &= 450 \text{ dynes/cm for mercury,} \end{aligned}$$

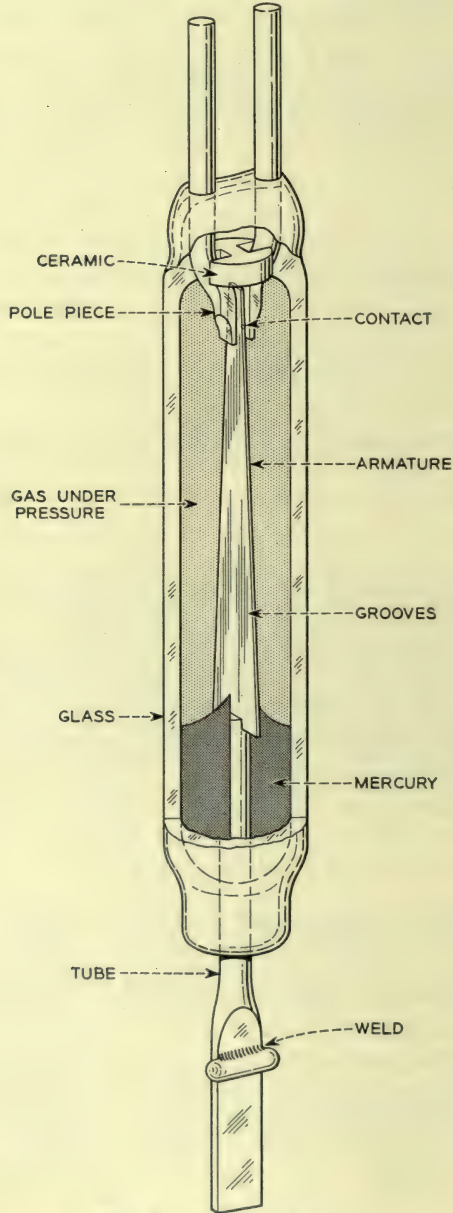


Fig. 2 — Vertical switch.

we obtain

$$h = 0.0338 \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \text{ cm.} \quad (3)$$

Fig. 3 shows a cross section of the armature and pole-piece at the contact, indicating the configuration of the mercury in the grooves and around a closed contact, as determined from equation (3). In the grooves, it will be noted, the surface is concave cylindrical, with a radius of -0.0163 cm, providing a path from the reservoir at this height with nearly the full capacity of the groove. The contact angles to the armature

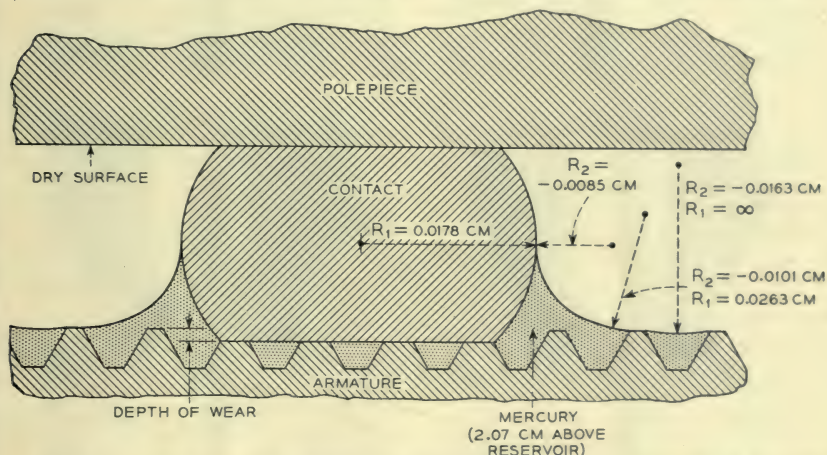


Fig. 3 — Cross-section showing configuration of mercury surface in grooves and around closed contact.

and platinum alloy contact surfaces are all zero, as these show complete wetting. The mercury fillet around the contact has positive curvature in the plane of the armature ($R_1 = +0.0178$ cm) and negative curvature ($R_2 = -0.0085$ cm) in planes through the axis of the contact.

The depth of wear indicated in Fig. 3 is brought to a stable value by an aging process in production.

DYNAMIC MERCURY CONFIGURATION

Fig. 4 shows flash photographs of the contacts at various instants during operation at 60 cps. It will be noted that, as the armature moves away from the fixed contact on the right, Figs. 4(a), (b) and (c), a bridge is drawn out which breaks in two places, leaving a free drop which falls out, Fig. 4(c).

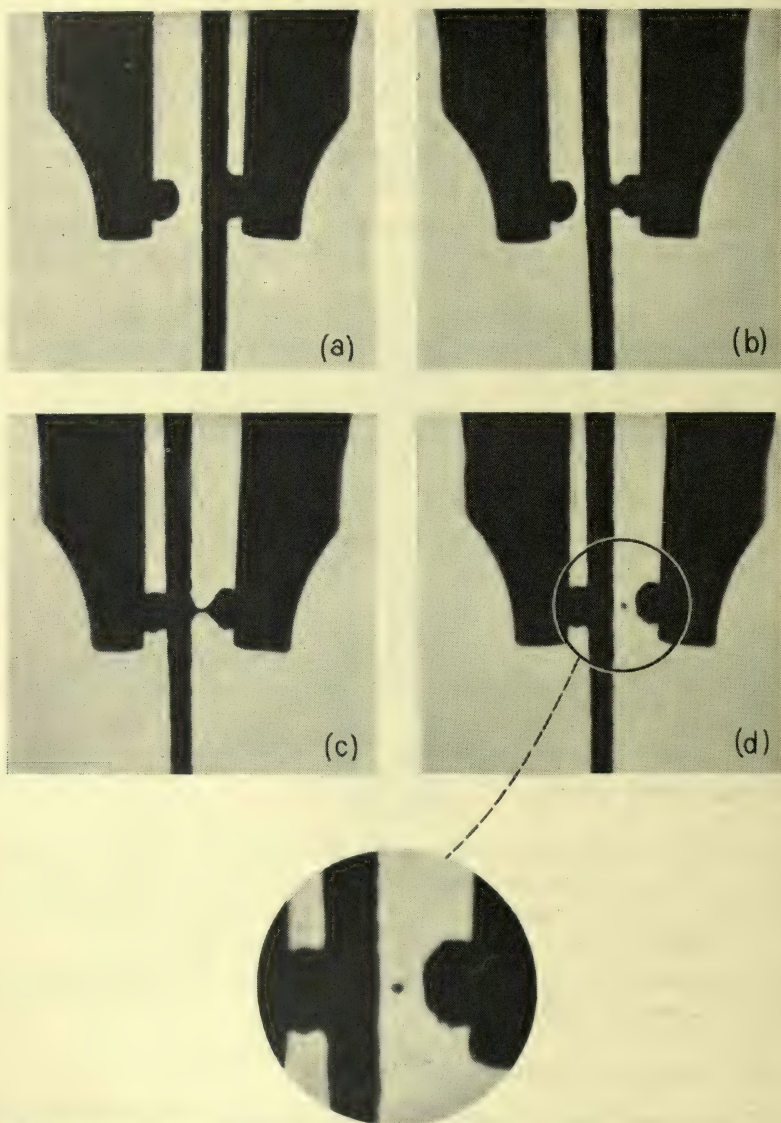


Fig. 4 — Dynamics of mercury contact. (a) Armature against contact on right. Note spherical shape of mercury on left contact. (b) Armature moves to left, drawing out mercury bridge. (c) Armature further to left; bridge about to break. (d) Armature reaches contact on left. Mercury bridge has broken in two places, leaving a tiny ball which later drops away.

When the ball falls out as the result of an operation, the loss of mercury results in a small, temporary constriction of the fillets in the grooves near the contact that was just opened. That is, the curvature of the surfaces of these fillets becomes more negative. As shown by equation (1), this produces a local decrease in liquid pressure. Mercury therefore flows up the armature from the reservoir to restore the normal static balance between surface tension and negative head. This is the fundamental behavior of an ordinary wick. The ball drops into the reservoir unless it happens to hit the armature on the way down. Thus, for repeated operations, there is a continuous circulation of mercury.

As indicated in Fig. 3, the fixed contact is a ball, with a flat surface where contact is made to the armature. In Fig. 4(c), taken directly after the breaking of the mercury bridge, the mercury remaining at the fixed contact on the right has been thrown back on the contact by surface tension forces, laying bare the flat surface. After several flow oscillations that are not shown, it comes to rest with the spherical contour shown by the contact at the left in Fig. 4(a). That is, being disconnected from the reservoir and having a limited wet surface to spread over, it assumes a positive head corresponding to a positive spherical radius about equal to that of the contact. This provides a mercury "cushion" in the form of a segment of a sphere, to which contact is made when the relay operates.

ADJUSTMENT OF SENSITIVITY

For various combinations of MMF's (magnetomotive forces) in the two magnets, various corresponding pairs of sensitivity values exist for operation in the two directions. A theoretical analysis of this relationship is given in the attached appendix. The adjustment of magnet strengths to obtain a specified pair of sensitivity values is made on the completed relay, using two electromagnets placed outside the can opposite the relay magnets. An automatic circuit is provided for this purpose that makes a complete adjustment in about 15 seconds, the time being dependent upon the precision required and the uniformity of the product being adjusted. The procedure used and the basis for it are discussed in the Appendix.

All of the adjustments used are of the type for which the armature moves all the way from one contact to the other when an operate current is applied. This represents the condition for the minimum differential ampere turns obtainable between the two operate values because it makes use of armature momentum.

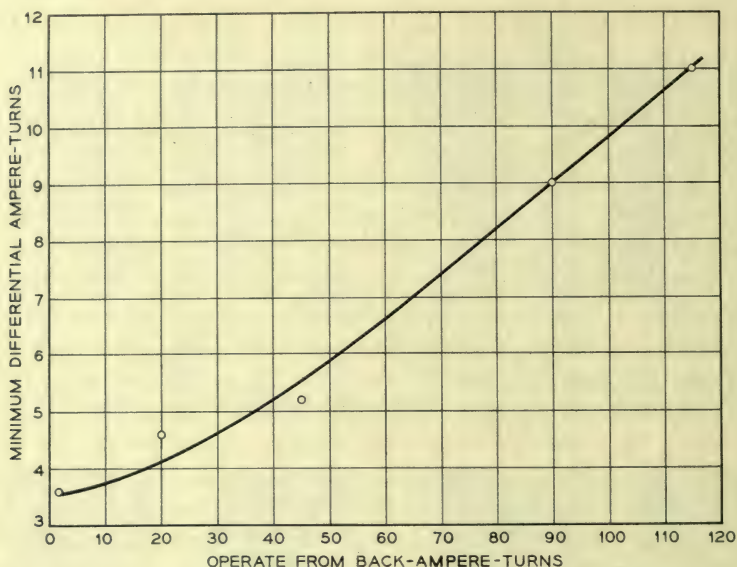


Fig. 5 — Minimum differential ampere turns as function of magnetic bias.

BIAS ADJUSTMENTS

As bias adjustments are made magnetically, rather than with the spring used in some polar designs, the armature flux tends to approach the same value at the operating points as it has in a balanced adjustment with the same differential ampere turn value. Such difference as does enter with increase of bias is due to differences in armature flux distribution between MMF introduced by the coil and that introduced by the magnets.

The effect of bias is illustrated in Fig. 5, which shows the minimum differential ampere turn value obtainable in a relay as a function of the higher operate value. The minimum differential value of 3.6 NI with a balanced adjustment is about the same as is obtainable with other polar relay designs. For operate values of about 100 ampere turns a "release to operate ratio" of about 90 per cent is obtained.

Larger differential ampere turn values, both with and without bias, are of course possible. The amount of bias obtainable in combination with larger differentials is somewhat reduced because of the greater magnet strength required.

EFFECT OF MAGNETIC SOAK ON SENSITIVITY

The effect of "soak" on sensitivity is illustrated in Fig. 6. It shows the change which takes place when the ampere turns to operate are

measured after applying a given ampere turn soak in each of two polarities.

SPEED OF OPERATION

Fig. 7 shows typical operate times of the relay for one balanced adjustment and one biased adjustment. The ordinates correspond to time to effect closure at the opposite contact after input is applied to the coil. The abscissae are shown both in terms of power in the full relay winding and the corresponding number of ampere turns. Two circuit conditions are indicated, one in which the voltage was applied directly to the full winding and the other in which the voltage was applied through a resistance equal to the coil resistance.

The winding used in this case is one which is specially designed for speed rather than sensitivity, being shorter in length, with the working gap near the middle. Its time constant (inductance-resistance ratio) is about 0.0006 second.

Fig. 8 shows "release" time measurements, where the relay, with various biased adjustments, was operated by opening the circuit. The curve is shown plotted against the "operate" setting of the relay. In this form it is relatively independent of the differential ampere turn adjustment.

The time required to open the contact from which the armature moves is typically about 0.0002 second longer than the closure time, as the

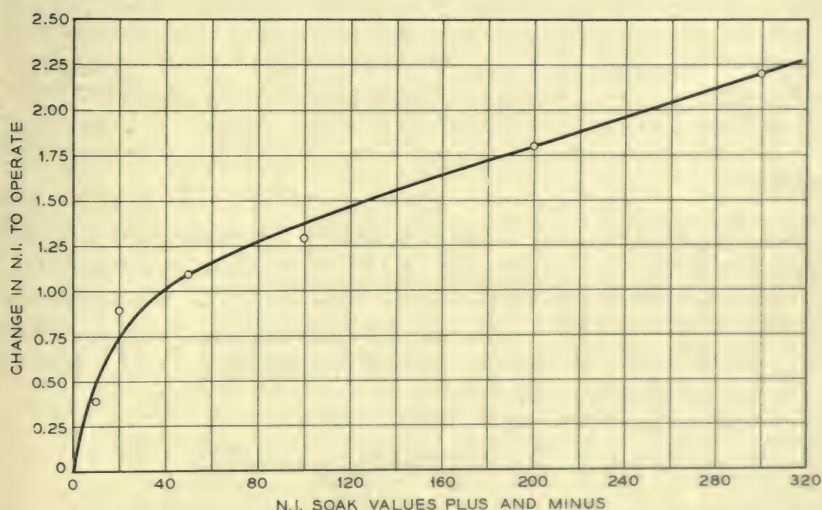


Fig. 6 — Change in N.I. To operate versus plus and minus NI soak.

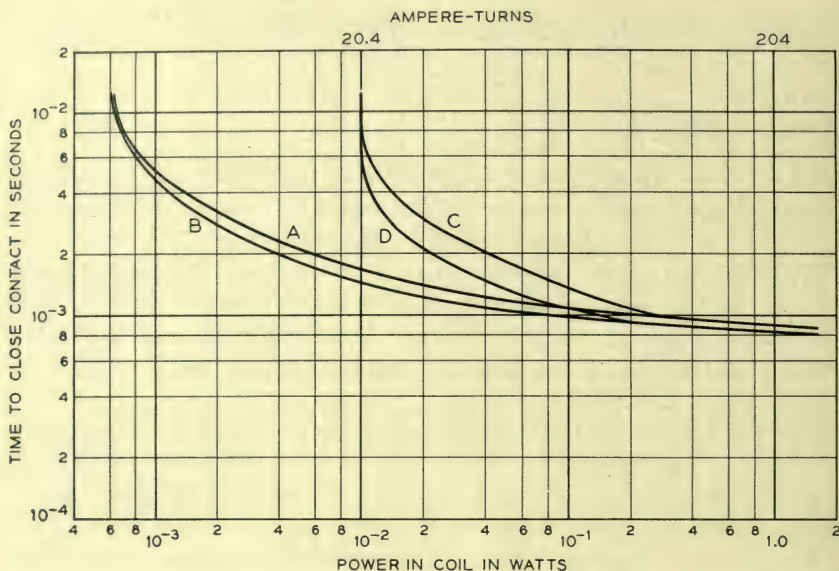


Fig. 7 — Time to close opposite contact after voltage is applied to coil. (A) ± 5 NI adjustment, voltage directly on coil. (B) Same as (A), except with resistance equal to coil in series. (C) Operate +20.4 NI, release +9.9 NI voltage directly on coil. (D) Same as (C), except with resistance equal to coil in series.

mercury bridge does not ordinarily break until after contact is made on the opposite side.

The natural frequency of the free armature, wet with mercury, is about 220 cps. The relays have been used at frequencies up to about 350 cps where the drive conditions were individually selected with respect to the phase of impact transients. Fairly well controlled operation without such selection can be obtained up to about 100 cps.

CONTACTS

The relay has been designed with telephone, rather than power applications in mind. The contacts are smaller than those in the previous neutral design, and this appears to be associated with less capability for closure of very high current circuits. The capacity required across the contact to prevent noticeable arcing appears to be about the same as that required in the earlier type:

$$C = \frac{I^2}{10} \text{ microfarads,}$$

but in most cases, larger values than this will be needed to hold peak

voltages to safe values. A small resistance in series with the condenser to limit the closure current is usually necessary. A considerable amount of very satisfactory experience has been had with inductive loads of 0.5 ampere at 50 volts, with 0.5 mf in series with 10 ohms across the contacts.

The contact closure shows no chatter for time intervals of 0.1 microsecond or more.

LIFE

Tests of relays with protected contacts indicate that the only change is a sensitivity change due to wear. Under conditions producing fairly high velocity of contact impact this change is of the order of 5 ampere turns increase in differential ampere turns for a billion operations, the change being roughly proportional to the logarithm of the number of operations.

HORIZONTAL TYPE RELAY

Most of the apparatus in the telephone plant is mounted on vertical panels. The substantially vertical mounting requirement for the relay shown in Fig. 1 tends therefore to be uneconomical from the space standpoint. Fig. 9 shows a modification of the glass enclosed switch which avoids this limitation by being designed to operate with its axis horizontal.

The switch modification consists in adding to the switch a special

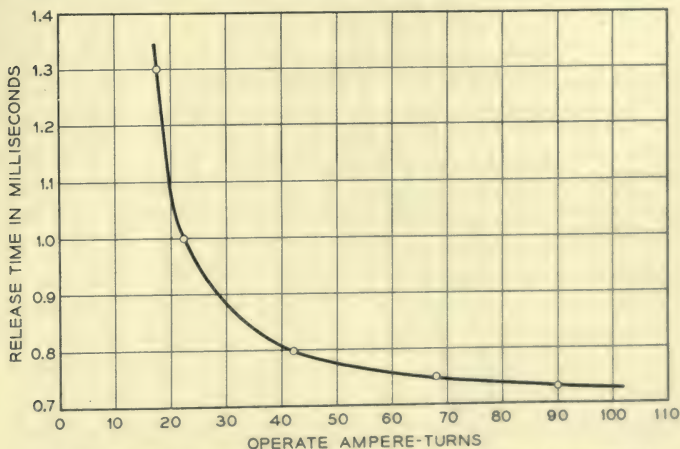


Fig. 8 — Release time versus operate ampere turn adjustment.

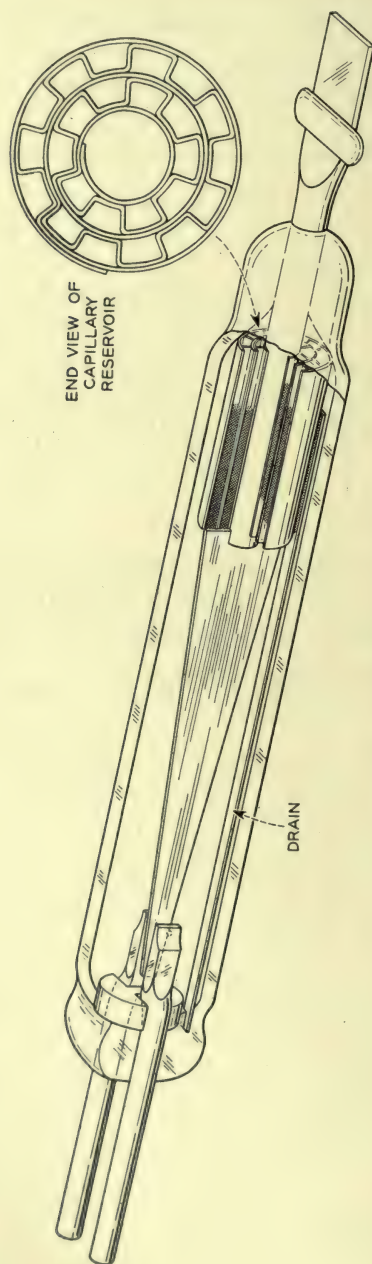


Fig. 9 — Horizontal switch.

capillary reservoir which establishes a negative head in the mercury equal to that obtained by the height of the contacts above the mercury reservoir in the switch shown in Fig. 2. To provide for the return of free mercury to the reservoir, a wet strip of metal is placed along the glass wall underneath the contacts.

The capillary reservoir is essentially a bundle of tubes with walls wet by the mercury. The amount of mercury in the switch is such that the tubes are about half full. The mercury menisci in these tubes are tangent to the walls and therefore have a spherical surface with a radius equal to that of the tubes. The proper tube radius ($R_1 = R_2 = 0.033$ cm) for

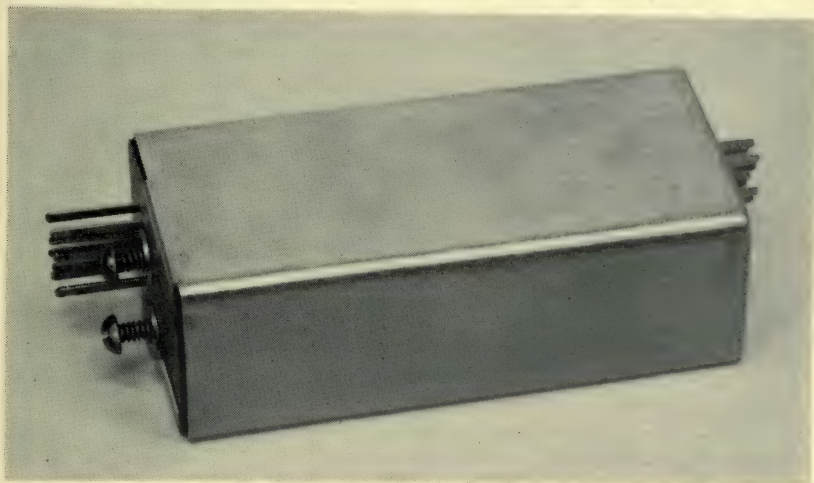


Fig. 10 — Horizontal relay.

the desired negative head ($h = 2.07$ cm) is obtained from equation (3). As the tubes have a uniform bore, the negative head which they establish is not critically dependent on the amount of mercury in them.

The capillary reservoir is made from thin strip, formed and then rolled into a cylinder. It surrounds the metal tube and is welded to the base of the armature. The drain element is sealed into the glass at one end, and, at the other end, is held in contact with the capillary element by the surface tension and negative head of the mercury.

The horizontal switch is an experimental design and has not been put into production. It can be assembled in the housing shown in Fig. 1 for plug connection to a vertical panel. Fig. 10 shows a preliminary model with the new type terminals for wrapped connections.² It is adapted for

² Solderless Wrapped Connections, B.S.T.J., May, 1953.

mounting on a vertical panel interchangeably with existing polar types. The terminals, which may be up to eight in number, are also brought out at the rear for test purposes.

HIGH-FREQUENCY SWITCHING

The small size, symmetry, and simplicity of the vertical switch structure has been found to be particularly well adapted for incorporation into coaxial structures for high-frequency switching and it is being used for a variety of applications of this kind.

USES

Small scale production of the relays has been started for a few special uses. Designs for larger scale uses are still in a preliminary stage. In general, it is expected that the relay will be well adapted for applications where a transfer contact in combination with low inductance, high speed, high sensitivity, low contact resistance, freedom from chatter, good high frequency switching characteristics, stability, long life, or any combination of these, is required.

APPENDIX

MAGNET STRENGTH VERSUS SENSITIVITY

A simplified representation of the magnetic system of a relay of this general type is shown in Fig. 11. Here the armature is shown working between two magnetic poles N and S , each of which has an area A . The armature position is indicated as a deviation ℓ toward N from the mid position, and is limited in its motion by stops at $\ell = \pm \ell_1$ from moving through the full magnetic gap range, defined by the positions $\ell = \pm \ell_2$.

M_s and M_n are MMF's in ampere turns, introduced by the magnets on either side, with positive values as indicated by the arrows. Similarly, M_c represents the MMF introduced by the operating coil. These values are not those of the magnets and coil in the actual relay. Instead they are assumed to be "Thevenin" equivalent open circuit values of MMF looking away from the working gaps on either side. The gaps are assumed to have the simple geometric dimensions of length $\ell_2 - \ell$ and $\ell_2 + \ell$, and area A . The reluctances looking away from the working gaps are assumed to be represented by the fixed gaps of area A and length $\ell_2 - \ell_1$ on either side.

The magnetic pull on the armature in a gap of this kind is

$$f = \frac{2\pi}{980} (0.1M)^2 \frac{A}{g^2} = \frac{(M)^2}{(124.9)} \frac{A}{g^2} \quad (1)$$

where f is the force in grams g is the gap length and M is the MMF in practical ampere turns across the gap.

In the structure shown, therefore, the net force on the armature from the gaps on both sides is

$$f_m = \frac{A}{(124.9)^2} \left[\frac{(M_n + M_c)^2}{(\ell_2 - \ell)^2} - \frac{(M_s - M_c)^2}{(\ell_2 + \ell)^2} \right], \quad (2)$$

where the values of M are expressed in ampere turns.

If $M_n = M_s$ this can be converted to

$$\frac{f_m}{f_{ml}} = \left[\frac{\left(1 + \frac{M_c}{M_n}\right)^2}{\left(1 - \frac{\ell}{\ell_2}\right)^2} - \frac{\left(1 - \frac{M_c}{M_n}\right)^2}{\left(1 + \frac{\ell}{\ell_2}\right)^2} \right], \quad (3)$$

where $f_{ml} = \left(\frac{M_n}{124.9}\right)^2 \frac{A}{\ell_2^2}$.

Fig. 12 shows curves of $\frac{f_m}{f_{ml}}$ versus $\frac{\ell}{\ell_2}$ for various values of $\frac{M_c}{M_n}$. Only one quadrant is shown as the system is symmetrical. It provides a convenient means for analysis of this type of relay.

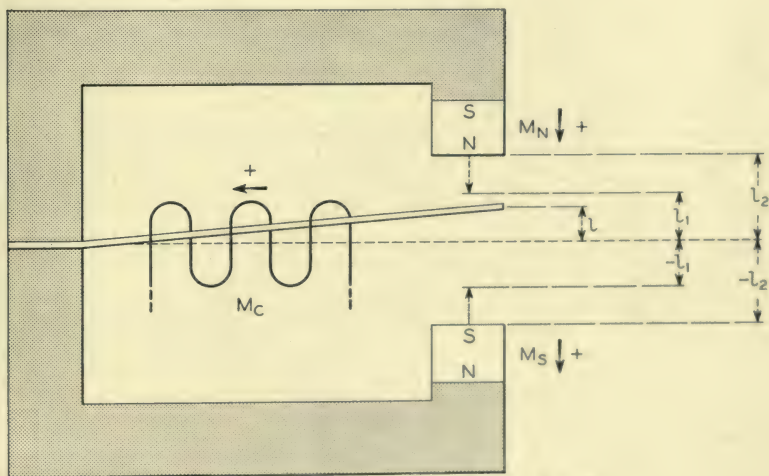


Fig. 11 — Simplified representation of relay.

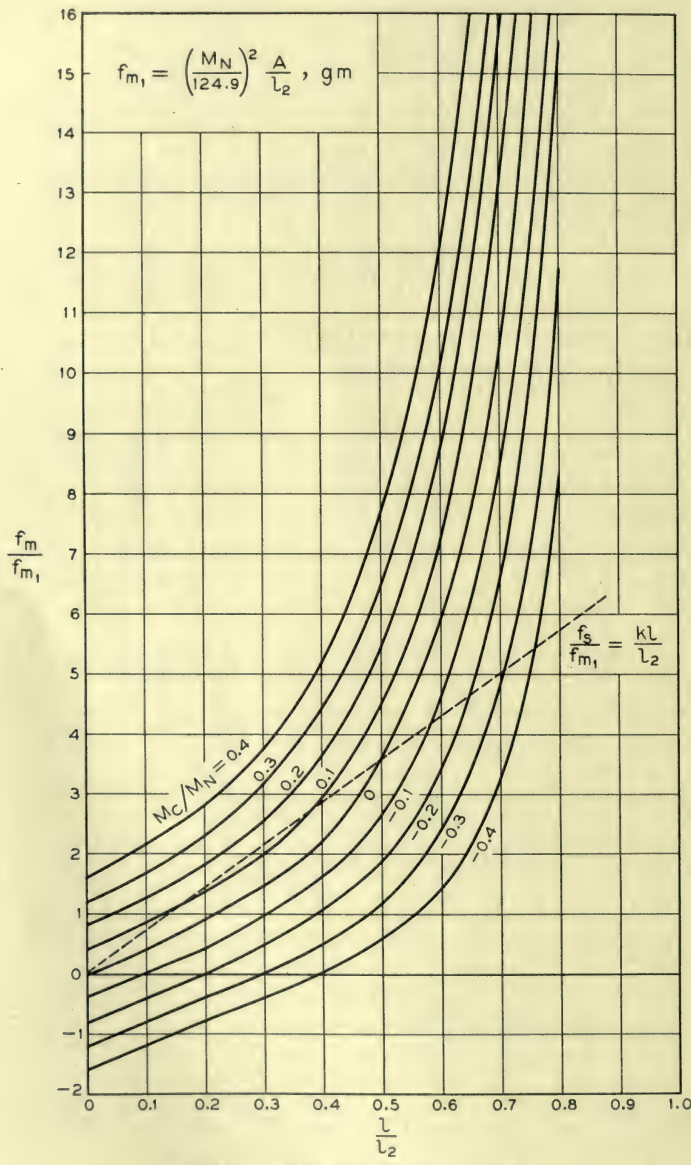


Fig. 12 — Balanced gap relay chart.

Also shown on the figure is a typical spring characteristic

$$\frac{f_s}{f_{ml}} = \frac{K\ell}{\ell_2} \quad (4)$$

If the stops ℓ_1 and ℓ_2 are placed at points of intersection of the spring characteristic with the magnetic characteristic for $M = 0$, (in this case at $\ell_s/\ell_2 = 0.48$), we have a condition that corresponds to a hypothetical relay of infinite sensitivity. That is, the armature can be released from contact with either side with an infinitesimal coil input, and, if there are no magnetic or mechanical losses in its travel, it will just swing to the opposite side without any change in the total energy of the system. Let us define the magnet strength for this condition as M_0 and let us assume that $M_c = 0$ for operation in either direction with this magnet strength.

Assume now that the magnet strengths on each side are increased equally to the values

$$M_{n1} = M_{s1} = M_0 + \Delta. \quad (5)$$

For values of ℓ_1/ℓ_2 near 1 this change can be balanced out by a coil input of about the same amount, as practically all of the pull on the armature would be from the nearer pole. For values of ℓ_1/ℓ_2 near 0 the effect of a coil input will be equal and aiding in both gaps. The coil inputs required to just operate from the N and S poles, defined as M_{cn1} and M_{cs1} , respectively for this particular case, will be

$$M_{cs1} = -M_{cn1} = p\Delta = p(M_{n1} - M_0) = p(M_{s1} - M_0), \quad (6)$$

where p is a value between 1 and 0.5. Values for individual cases can be worked out with reference to the curves for various values of M_c/M_n in Fig. 12. For the case shown, where $\ell_1/\ell_2 = 0.48$, p is about 0.8.

An adjustment in accordance with (6) is thus a balanced one with a spread of $2p\Delta$ between the two sensitivity values, the amount of spread increasing with increase in the strength of the two equal magnets M_{n1} and M_{s1} above the value M_0 .

A general type of adjustment, including all possible combinations of the two sensitivity values, can be obtained by adding a suitable value B to a balanced pair of sensitivity values in accordance with equation (6). These general sensitivity values would then be

$$\begin{aligned} M_{cs} &= M_{cs1} + B, \\ M_{cn} &= M_{cn1} + B. \end{aligned} \quad (7)$$

This is the type of change that would be produced by a bias of $-B$

ampere turns in an auxiliary coil. The equivalent of such a bias would be obtained by decreasing the strength of the N magnet and increasing the strength of the S magnet, each by the value B . The general magnet strength values would then be

$$\begin{aligned} M_n &= M_{n1} - B, \\ M_s &= M_{s1} + B. \end{aligned} \quad (8)$$

Combining equations (5), (6), (7) and (8) to eliminate M_{cs1} , M_{cn1} , M_{n1} , M_{s1} and B , we obtain the general relations between sensitivity and magnet strength as

$$M_{cs} = p \left(\frac{M_n + M_s}{2} - M_0 \right) + \frac{M_s - M_n}{2}, \quad (9)$$

$$M_{sn} = -p \left(\frac{M_n + M_s}{2} - M_0 \right) + \frac{M_s - M_n}{2}. \quad (10)$$

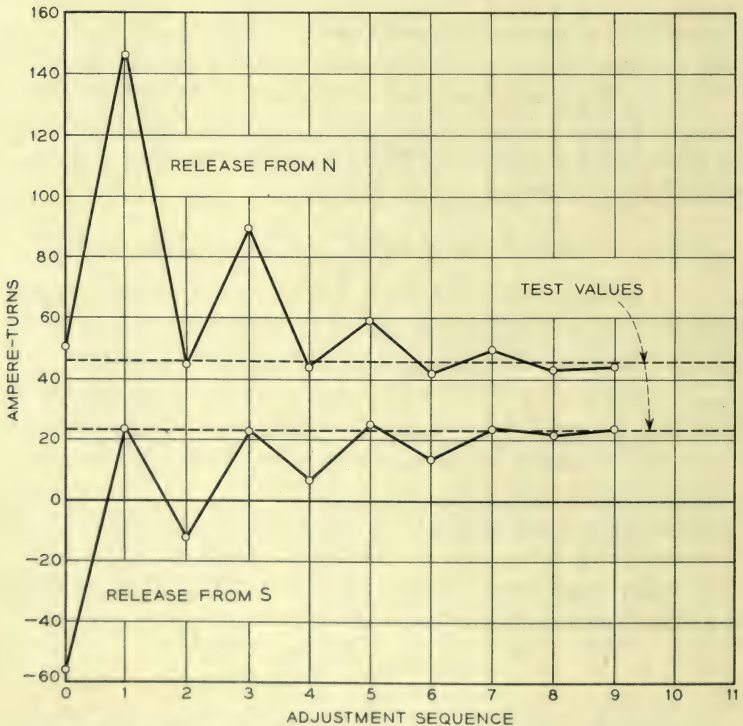


Fig.13 — Sequence of sensitivity values obtained in adjustment.

ADJUSTMENT OF SENSITIVITY

The procedure used to adjust the relay to any desired pair of sensitivity values is as follows:

1. Magnetize the two magnets fully by a flux directly across them.
2. With the armature starting from one side, progressively decrease the magnetization on that side in small steps until the armature just operates to the other side on the final current value desired for this direction of operation.
3. With the armature starting on the other side, progressively decrease the magnetization on that side in small steps until the armature just operates from that side on the final current value desired for this second direction of operation.
4. Repeat 2 and 3 alternately until the relay just operates in both directions on the two current values desired.

Within limitations such as the size of demagnetizing increments used per step, this procedure results in an adjustment to the two test values used. Let us consider why this result is obtained.

If we should substitute $p = 1$ in equations 9 and 10 we would obtain

$$\begin{aligned}M_{cs} &= M_s - M_0, \\M_{cn} &= M_1 - M_n.\end{aligned}$$

That is, in such a case, the two adjustments would be independent of each other and the above process would require only one adjustment of each magnet. In the more general case, where the force on the armature is affected by both poles, each adjustment on one side results in a greater magnet strength at the pole being adjusted than is required for the final adjustment, because of the pull from the opposite pole, the amount of which is greater than normal because that pole has not been reduced to its final magnet strength. Thus the final adjustment is normally reached through a number of alternations between the two sides, which progressively approach the final pair of sensitivity values.

The sequence of sensitivity values obtained in a typical adjustment of this kind is shown in Fig. 13.

Dynamic Measurements on Electromagnetic Devices

By M. A. LOGAN

(Manuscript received July 6, 1953)

A sampling switch with adjustable make and fixed break times can be used to obtain dynamic measurements of reciprocating phenomena. A test set has been developed using this principle to measure the flux-time, current-time, displacement-time, and velocity-time response of electromagnets and similar devices. The tested device is steadily cycled. A dc instrument is switched in by the synchronous control at any preselected instant in the cycle, and out when a steady reference condition has been reached. The steady reading of the instrument is proportional to the value, at the closure instant, of the variable being measured. The instrument switching is controlled by an electronic timing system. This system operates mercury contact relays, which do the actual switching. For the displacement-time and velocity-time measurements, an optical transducer with associated dc amplifiers is added. The design of these devices is described. The results of an investigation of dynamic flux rise and decay in solid core electromagnets are reported. Modified first approximation equations are developed to give a better representation of eddy current effects.

INTRODUCTION

The application of common control methods to telephone switching systems has led to the widespread use of high-speed relays. The actuation time of these relays is affected by many parameters such as the power supplied, how far the armature has to move, the mechanical work the armature has to perform during its motion, the winding design, the magnetic structure, and eddy currents introduced in the magnetic members caused by the application of current to the winding. The eddy currents act to oppose a magnetic flux change and hence retard a building or decay of flux. This causes the actuation time to be increased compared to a relay without such effects. An analytical determination of the development and effects of eddy currents can be made for simple sym-

metrical magnetic structures having an infinitely long or torus shaped round or rectangular cross-section, assuming linear magnetization characteristics. However, for relay-like structures having air gaps, leakage flux which only partly completes its circuit through the magnetic material, varying cross-section so that boundary conditions become complicated, and non-linear magnetic properties, an analytical approach becomes unmanageable.

For a fundamental study and direct measurement of eddy current effects, a test set has been developed to measure the dynamic flux rise and decay characteristics of relays, and similar structures. This test set is electronically operated on a synchronous switching principle. It displays on an ordinary dc instrument, as a steady reading, the instantaneous flux obtaining at any selected time after energization or de-energization of the magnet.

The application of the test set is restricted to devices which operate under conditions of suddenly applied or removed dc voltage and which can be cycled between these two conditions until the dc instrument reaches its steady state reading.

This article will present the theory of the basic synchronous switching circuit and the relation applying between the dc instrument reading and the instantaneous flux linkages, a description of the electronic control circuits, and measurements of dynamic flux rise and decay using solid core electromagnets.

Supplementary additions to the basic circuit are available using the same principles, to measure current-time, displacement-time, and velocity-time curves of reciprocating devices. The first fundamental measuring set using the switching principle, was built by E. L. Norton¹ in 1938. The earlier set used a synchronous, motor driven, phase adjustable commutator, to perform the switching. A limitation of the earlier set in how fast it could be operated, combined with brush wear and chatter troubles led to the development of the new electronically controlled set employing sealed mercury contact relays to perform the switching.

Part I—Description of Fluxmeter System

BASIC MEASURING CIRCUIT

A schematic of the basic measuring circuit is shown in Fig. 1. A battery switching contact indicated as A , applies and removes voltage to the magnetizing winding of the electromagnetic device under test with a 50 per cent duty cycle. The time of one complete cycle is indicated as T . The electromagnetic device is equipped with a search coil of N turns, having dc resistance, including wiring, of R_e ohms. A B contact, synchronously switched at instants to be described later, connects a dc microammeter to the search coil. A damping contact C connects the instrument to a resistance, preset to the same value as the search coil, when the instrument is not connected to the latter, thus providing the same instrument damping at all times.

The A timing cycle is shown schematically for one interval T . Next below is shown the cycle for the B contact, followed by that for the C contact. The B contact always opens and the C contact closes just before the A contact closes. The point of closure of the B contact, indicated as t_1 may be adjusted to occur at any time during the cycle T . The cycle time T is chosen sufficiently long so that the flux in the electromagnet has sufficient time to reach a substantially steady state during both the closed and the open intervals.

The flux rise and decay have the general characteristics shown as the φ curve, while below it is the induced voltage in the search coil, exactly given by Lenz's Law

$$e = N \frac{d\varphi}{dt}. \quad (1)$$

The first property of this last curve to observe is that the positive and negative areas are exactly equal. Otherwise a dc instrument connected permanently to the search coil would show a dc current flow. The second property is that the area of the curve up to some time t_1 is proportional to the instantaneous flux at the time t_1 . Finally, this unshaded area is numerically equal but opposite in sign to the shaded remainder of the area, taking into account the signs of the two components of the remainder. The operation of the test set depends upon measuring this shaded area.

Let the flux linking the search coil of N turns at any time be φ . Assume that the microammeter has a resistance R_m and inductance L_m . The current flowing in the meter mesh while the B contact is closed is

$$N \frac{d\varphi}{dt} + L_m \frac{di}{dt} + (R_m + R_e)i = 0. \quad (2)$$

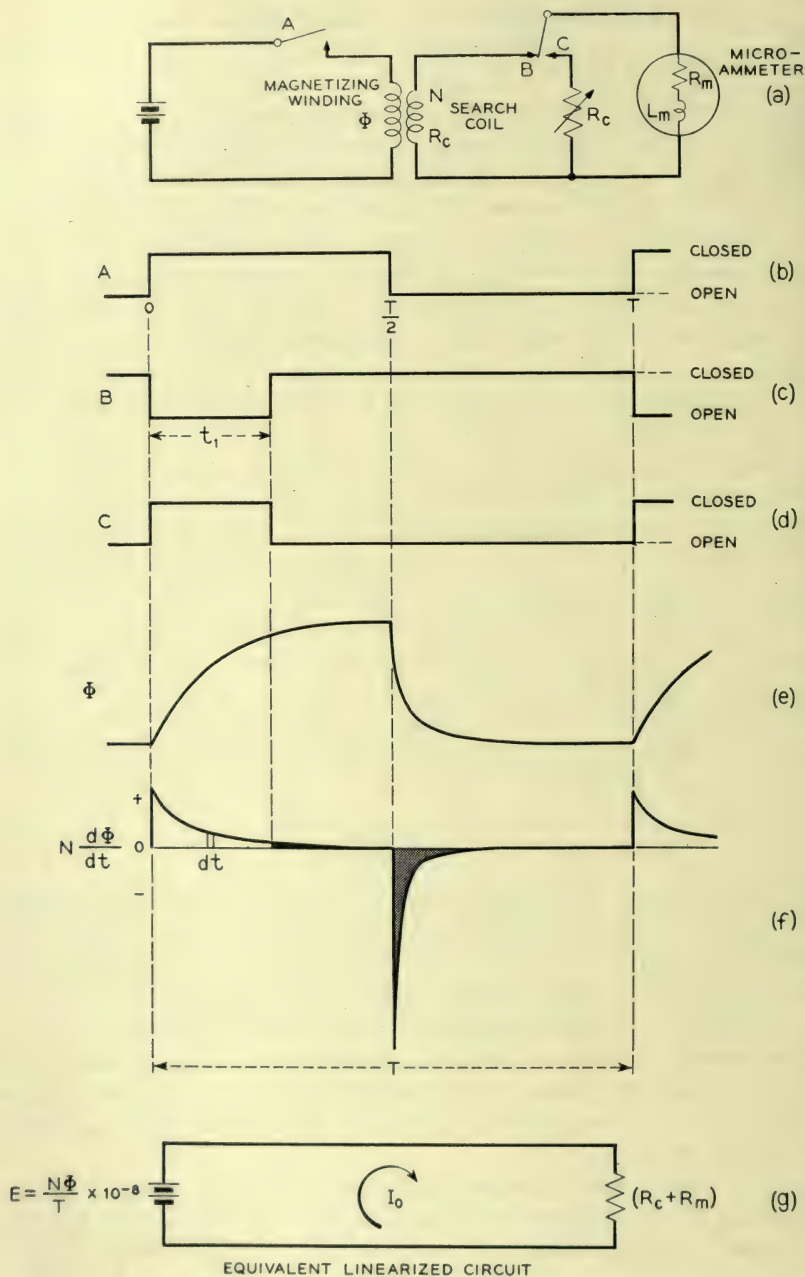


Fig. 1 — Fundamental operation of dynamic measuring set.

The dc instrument will read the average value of current for the cycle. As the period of the cycle is T , and t_1 is the chosen delay for closure of the B contact measured from the time of the A closure, the direct current indicated by the instrument will be:

$$I_0 = \frac{1}{T} \int_{t_1}^T i \, dt = -\frac{1}{(R_m + R_c)T} \int_{t_1}^T \frac{d}{dt} (N\phi + L_m i) \, dt$$

$$= \frac{(N\phi + L_m i)_{t_1} - (N\phi + L_m i)_T}{(R_m + R_c)T} \quad (3)$$

Now the product $L_m i$ at t_1 when the B contact is closed must be zero. Furthermore, if the period T is long in comparison with the time of rise and decay of the flux, and with the time constant L/R of the measuring circuit, the product $L_m i$ is also negligible at $t = T$. We then have:

$$\Phi_{t_1} - \Phi_T = \frac{RTI_0}{N} \times 10^8 \text{ Maxwells}, \quad (4)$$

where R is now the total resistance of the measuring circuit. The factor 10^8 has been added to convert from practical units to c.g.s. units. The flux Φ_T is the residual flux, so that if this is taken as a reference value, the value of the flux at the time of closing the B contact is simply

$$\Phi = \frac{RTI_0}{N} \times 10^8 \text{ Maxwells}. \quad (5)$$

It may be objected that the flux in equation (2) is made up not only of the flux to be measured but also of a component due to the current in the measuring circuit. This is true, and the flux linking the search coil changes differently during the time the B contact is closed compared to that without the instrument circuit. Note, however, that we are not concerned with how the flux varies between the times t_1 and T but only with its value at the limits. Since the flux at t_1 is that which has been established with the search coil circuit open, the one requirement is that the constants be such that the current in the measuring circuit be substantially zero just before T . This can be verified through the measurements themselves by noting whether there is any change in measured flux in this interval. If there is, then the cycle time T can be increased until this requirement is met.

It will be convenient later to regard the switched meter as a linearized circuit composed of an equivalent dc voltage

$$E = \frac{N\Phi}{T} \times 10^{-8} \text{ volts} \quad (6)$$

in series with the resistor R , causing a dc current I_0 to flow, Fig. 1 (g).

MEASUREMENT OF CURRENT

The instantaneous value of current in a circuit being cycled by the *A* contact is measured by inserting the primary of a toroidal air core transformer in series with the circuit as shown in Fig. 2. The secondary is connected to the meter through the *B* contact. The best design consists of a secondary occupying half the winding volume and wound to the same resistance as the meter. The primary can be wound in two sections half inside and half outside of the secondary to provide maximum mutual inductance, of a wire coarse enough to provide the number of turns for a convenient meter deflection. In an air core transformer, the magnetic circuit is linear and the flux exactly proportional to the current, a situation that does not exist when magnetic materials are present. Because of this linearity, the analysis can be made in terms of inductances and currents rather than flux linkages and currents.

By definition, the voltage induced in the secondary of a linear transformer is related to the primary current, through the mutual inductance *M* by the relation:

$$e = M \frac{di}{dt}, \quad (7)$$

which corresponds exactly to (1) with *M* substituted for *N*, and *i* for ϕ . Hence, the instantaneous current at time *t*₁ is given by the relation

$$i = \frac{RTI_0}{M}, \quad (8)$$

where the other symbols have the same significance as before. This relationship makes the calibration of the mutual inductance easy by setting up a battery and resistor for the primary circuit and measuring the steady state current with an ammeter. Then when the *A* contact is being switched and the instrument current *I*₀ is read, corresponding to the known final primary current *i*, relation (8) can be solved for *M*.

MEASUREMENT OF FLUX USING A BRIDGE CIRCUIT

The search coil method of measurement just described is preferred because of the absence of any dc voltage in the meter circuit. However

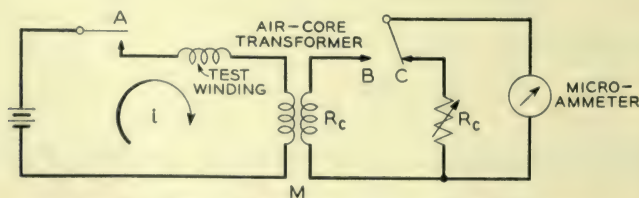


Fig. 2 — Circuit for the measurement of current.

sometimes the use of a search coil is inconvenient, or for other reasons it is desired to measure the flux linkages in the magnetizing winding itself. The test set includes a bridge circuit shown in Fig. 3 for such measurements. With the contacts *A* and *B* closed and the bridge in the flux rise arrangement, the bridge is first balanced for no current in the instrument, by adjustment of R_d . The same instrument is used for setting the balance as later used for the flux measurement. The damping resistance is next set to the value facing the instrument, namely,

$$R_c = \frac{K}{K+1} (R_0 + R_d). \quad (9)$$

If then the contacts are operated with the period T , the instrument will read a current I_0 , and the flux-turns at time t_1 during rise will be

$$N\Phi \times 10^{-8} = R_m I_0 T \left[1 + K \left(1 + \frac{R_0}{R_m} + \frac{R_d}{R_m} \right) \right] - L_1 i, \quad (10)$$

where N is the number of turns of the winding, Φ is the average flux per turn, and the other quantities are indicated on the drawing. The term $L_1 i$ is a correction term, usually negligible, involving the self inductance of the primary of the air core current transformer and i the instantaneous

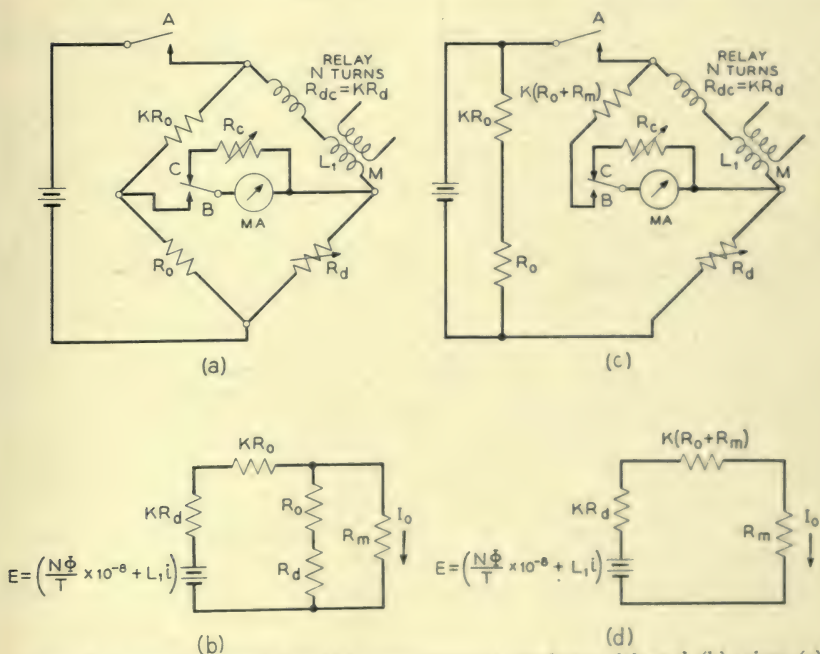


Fig. 3 — Bridge circuit for the measurement of flux — (a) and (b), rise; (c) and (d), decay.

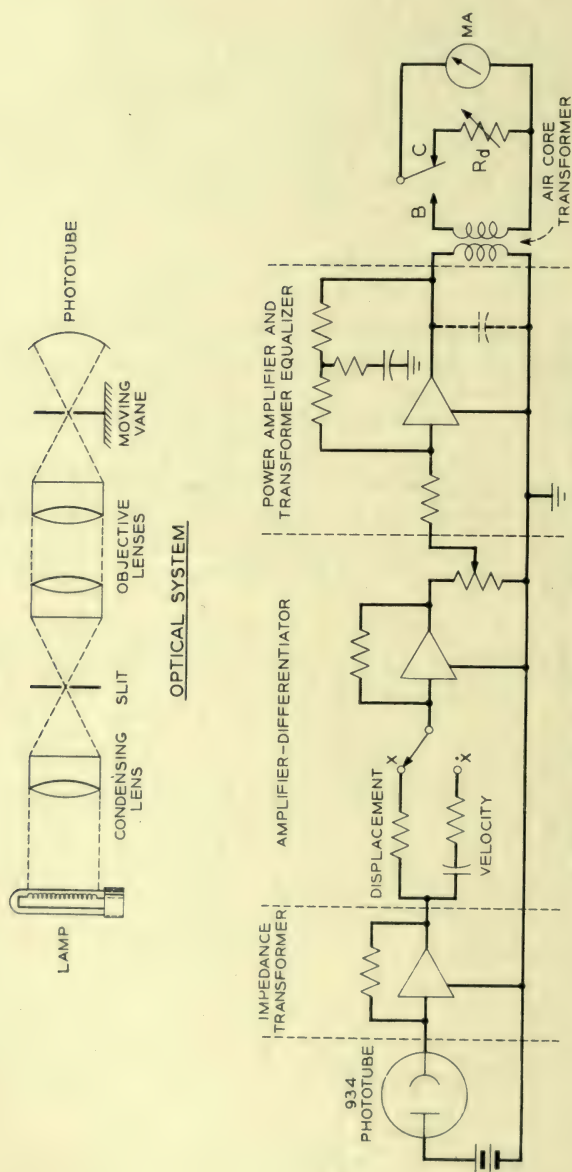


Fig. 4 — Block diagram of photocell amplifier system.

current at the time t_1 ; i is determined by a separate measurement described above. If its value is not needed, the air core coil can be omitted from the circuit.

For flux decay measurements made after the A contact has opened, the bridge circuit has to be modified to remove the resistance R_0 and KR_0 from being across the electromagnet and affecting its flux decay. They are transferred to the battery side of the A contact as shown in Fig. 3(c), to maintain the same battery drain. This avoids any error due to the internal resistance of the battery, which would otherwise cause the final flux at the end of a rise test to differ from the initial flux at the beginning of a decay test. The added resistor $K(R_0 + R_m)$ is connected in series with the meter to make the expression for the flux-turn linkages the same as before. For decay measurements the damping resistor is set at the value

$$R_c = K(R_0 + R_m + R_d). \quad (11)$$

In Figs. 3(b) and 3(d) are shown the linearized equivalent circuits for rise and decay respectively, from which equation (10) above can be derived conveniently.

MEASUREMENT OF DISPLACEMENT AND VELOCITY

An optical probe is provided, in which the amount of light falling on a photocell is controlled by the relative position of two flags, one cemented on each of the relatively moving parts to be studied, such as an electromagnet. The change in output current from the photocell thus is proportional to the displacement of the armature with respect to the core, one flag being on each. A block diagram of the system is shown in Fig. 4. Amplifiers effective from dc up to a frequency determined by the resolution required, with substantially no phase shift, deliver a current into the primary of an air core transformer proportional to the instantaneous displacement of the armature. By virtue of the linearity of this transformer, the flux developed is proportional to the displacement. Thus by operating the electromagnet with the A contact and connecting the secondary of the transformer to the B contact, the instrument gives a dc reading proportional to the armature displacement at the time the B contact closes. The total displacement can be measured statically. The instrument reading at a time after complete operation corresponds to this known displacement, permitting the scale to be converted to a displacement scale.

By changing one amplifier input resistor to a capacitor, the amplifier

becomes a differentiator, in which the output current is proportional to rate of change of input voltage. With this one change, the instrument reads the instantaneous velocity of the armature. Under these conditions the second amplifier differentiates the input voltage once, the output transformer differentiates a second time, and the dc instrument integrates once, leaving a net result of one differentiation.

DESCRIPTION OF SYSTEM COMPONENTS

FLUXMETER

General

In the following descriptions of the several components of the system the general characteristics required are considered, the specific deviations from ideal are determined, and an evaluation of the measurement errors is made. The description starts with the dc instrument followed by the associated vacuum tube filter. Then the heart of the system, the contact switching circuit itself is considered. Following this is the timing impulse generating circuit, the counting rings, the time selector, coincidence circuits, memory, and relay control circuits. These elements make up the fluxmeter proper.

The auxiliary circuits for displacement-time and velocity-time measurements are the concern of the next part of the paper. The components are the optical system, the photocell amplifier, the amplifier-differentiator and finally the output amplifier.

The last section first shows typical measurements on a telephone type relay. Then a description of a more fundamental study of dynamic flux rise and decay in solid core electromagnets is given. This study has led to two new first approximation equations for dynamic flux rise and decay as will be seen.

Dc Instrument and Effect of Damping Resistance

The dc instrument used for a majority of the measurements is a heavily damped 50.0-ohm, 200-microampere full scale instrument, with a $\frac{1}{2}$ per cent of full scale accuracy. The open circuit decay time constant is about 2 seconds, and with a shorted winding about 4 seconds. On Figure 5 is shown a plot of the decay time constant referred to a short circuit versus the damping resistance referred to the meter resistance.

An evaluation of the small error introduced by not providing the damping resistor is as follows. Consider a square wave flux pattern produced by an ideal electromagnet, with zero winding time constant or

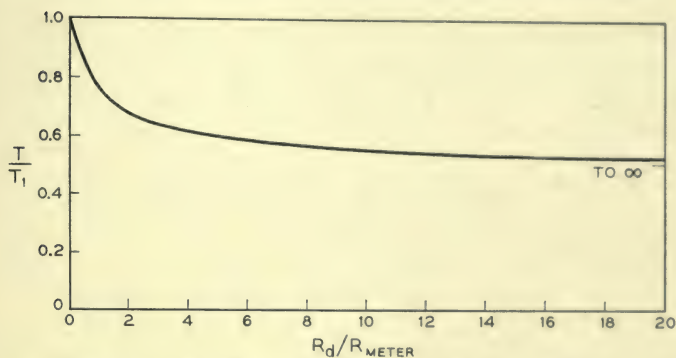


Fig. 5 — Decay time constant of DC instrument.

eddy currents. The induced voltage in the search coil then consists of two equal and opposite impulses occurring respectively at 0 and $T/2$. To measure the constant flux, the switching epoch can be anywhere between these two values. The switching omits the impulse at zero time. For limiting cases t_1 can approach either 0 or $T/2$. The instrument reading of course should be independent of the choice, as the flux remains at the constant maximum throughout this interval. The instrument receives only the decay impulse, and thus there will be an average dc component.

For the case of t_1 approaching zero, the instrument is always connected to the search coil, which usually is of negligible resistance and the short circuit meter decay time constant applies. Just before a succeeding impulse, the meter will have decayed to a relative value of:

$$e^{-T/T_1} \approx 1 - \frac{T}{T_1}, \quad (12)$$

where T is the cycle time as before and T_1 is the short circuit decay time constant. The pointer next abruptly rises to its original maximum value, followed again by another equal decay in a cyclic manner.

For the case of t_1 approaching $T/2$, the instrument is connected to the search coil only half the time, being open circuited for the remainder of the time, if no damping resistor is provided. It thus decays at two different rates following the impulse. The first half is the same as before but the second half is under the condition of open circuit. For small errors the relative decay will be:

$$e^{-T/2T_1} e^{-T/2T_2} \approx 1 - \frac{T}{2} \left[\frac{1}{T_1} + \frac{1}{T_2} \right], \quad (13)$$

where T_2 is the open circuit decay time constant.

The search coil impulse is the same in each case, and for small errors, the instrument will rise the same amount for either condition followed by equal decays. The two maximum values M_1 and M_2 are not equal but are related by:

$$M_1 \frac{T}{T_1} \approx M_2 \frac{T}{2} \left(\frac{1}{T_1} + \frac{1}{T_2} \right) \quad (14)$$

$$\frac{M_1}{M_2} \approx \frac{1}{2} \left(1 + \frac{T_1}{T_2} \right) \quad (15)$$

The ratio of the two instrument readings thus is a function of the ratio of the two instrument decay time constants. For the case of $T_1 = 4$ sec., $T_2 = 2$ sec., the limiting error is 33 per cent. This shows that without a damping resistor, the error can be many times the $\frac{1}{2}$ per cent instrument error, but that the timing for switching the meter damping is not critical. That is, small time gaps with the damper off will introduce an insignificant error. In a sense, the omission of the damper is equivalent to using an electronic integrator with finite internal gain.

Vacuum Tube Filter Circuit¹

The above discussion brings out the fact that visible motion of the instrument pointer without a filter occurs, increasing as the cycle time is lengthened. From equation (12) and using $T = 0.1$ sec., and $T_1 = 4$ sec.,⁸ the amplitude would be about $2\frac{1}{2}$ per cent of full scale. Reading the center of the vibration is easy for cycle times T less than 0.1 second and for such measurements the filter is not used. Many measurements, however, require a slow pulse rate and even a sluggish instrument will follow the pulses to such an extent that an accurate reading is impossible.

Fig. 6 is a schematic of a vacuum tube circuit which serves as a very efficient low pass filter. This filter must fulfill a variety of very stringent requirements, chief of which may be listed:

- (a) Low loss to direct current.
- (b) High loss, probably exceeding 30 db to frequencies above one cycle per second.
- (c) A constant resistance input.
- (d) Ability to suppress peaks of relatively high voltage of the order of several hundred or thousand times the dc voltage being measured.

(e) No effect other than one which may be accurately calculated on the dc reading.

(f) The circuit should not greatly increase the time required for the instrument to reach a steady reading.

Requirement (c) may call for some comment. It was pointed out earlier that the method requires the time constant of the measuring circuit to be small in comparison with the period of the pulses. A filter built in the ordinary way to have high suppression to pulses of period T would not have a time constant small in comparison with this figure. The require-

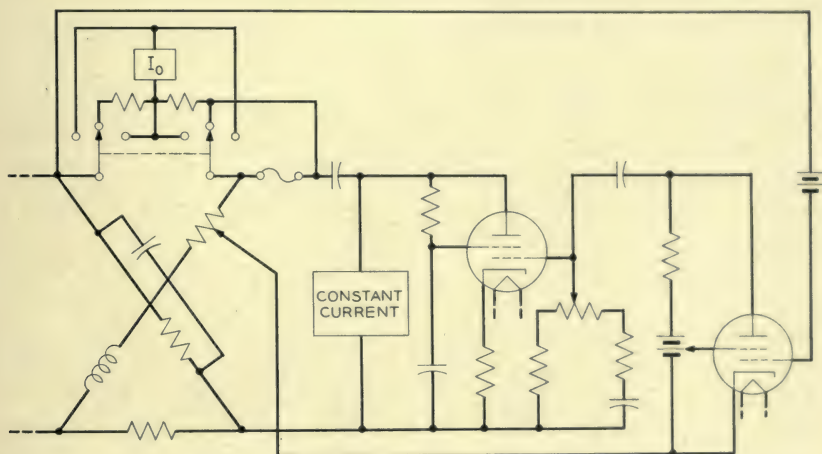


Fig. 6 — Vacuum tube low pass filter circuit.

ment of constant resistance input is a convenient way of expressing the necessity of fulfilling this condition.

The original work on the system used was done by R. F. Wick, and the features to be described are due to him and E. L. Norton. From left to right in Fig. 6, the elements are as follows: a balanced impedance bridge containing the ammeter in one arm, a blocking condenser, a constant current high impedance power supply for the power tube, and a two stage amplifier with an interstage phase adjusting circuit. The three-point double-pole key when in the normal position removes the meter (50 ohms) and substitutes a 50-ohm resistor. When off-normal, the meter is connected in either polarity.

The operation of the circuit is best understood by assuming the reactance elements to be omitted from the bridge and the contact on the potentiometer forming one diagonal to be at the lower left. The input to the amplifier is then directly across the line and any feedback is elimi-

nated, since the bridge is balanced. Any alternating component applied will be amplified and sent back through the meter in the opposite polarity to that coming directly from the line. If then the gain in the amplifier is correct and its phase shift nearly zero, the alternating component through the meter will be greatly reduced.

The bridge provides the constant resistance feature at the input, since the output of the amplifier can have no effect on the input impedance. It does however necessitate a loss from both line and amplifier to the measuring instrument.

Practically, the difficulty in design is in securing an amplifier with enough peak output capacity and negligible phase shift at frequencies from one cycle up. This has been accomplished by several methods. Most of the phase shift is introduced by the plate condenser in the last stage. The effect of this is reduced by the constant current power supply.

The gain of the amplifier is controlled by feedback secured by the potentiometer in the diagonal arm of the bridge. The gain is sufficient so that a substantial amount of feedback may be used with a consequent further reduction in phase shift.

The final phase compensation is secured by the interstage potentiometer. The effect of this is illustrated in Fig. 7. The lower curve is the net phase shift of the amplifier without the interstage circuit. The upper curve is the phase shift which may be introduced at the grid of the second stage. By proper adjustment, these may be made to compensate each other down to a frequency of one or two cycles. The circuit constants are such that the final adjustment of low frequency phase may be made with a negligible effect on high-frequency gain. It may be of interest that if the bridge is unbalanced by shorting the lower 50-ohm resistance, an error of 20 per cent or more is introduced in the dc reading due to the extremely large effective reactance of the feedback amplifier. If the bridge is unbalanced by shorting the instrument the amplifier is quite likely to "motor-boat" and blow the fuse used for protection. It is for this reason that the instrument key may be set to replace the instrument by a 50-ohm resistor. When in this position the fuse is also shorted to avoid needlessly blowing it due to the high surges which frequently occur when the amplifier is turned on and the condensers start to charge.

The filter as described, omitting the reactance elements from the bridge, would be entirely satisfactory within the power capacity of the amplifier. With certain measurements, notably those of velocity, the peaks of the wave as applied to the filter, which in this case would be proportional to acceleration, are so high that no reasonable amplifier would be able to handle them. These knifelike peaks however are so sharp that they may

easily be removed by reactance elements added to the bridge as shown. It will be noted that the bridge is still of constant resistance, and is still balanced at all frequencies.

Three adjustments are provided on the amplifier: one controls the feedback, a second the phase correction, and the third the plate current. The amount of suppression is controlled by a three position key which may be used to cut down the sizes of the capacitors in the circuit in a ratio of two and four to one. In cases where one of the higher pulse rates is used, adequate suppression and somewhat faster readings may be secured by using smaller capacitors. With the maximum suppression, two cycles may be reduced to such an extent that it has no detectable

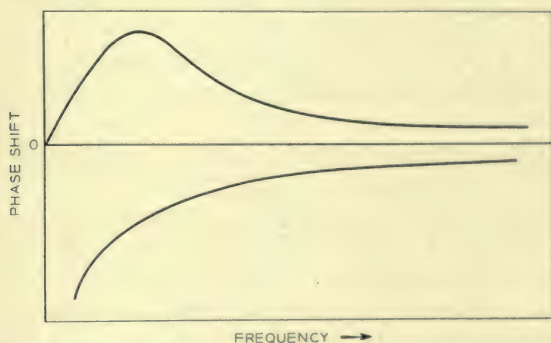


Fig. 7 — Illustration of method for phase compensation of amplifier.

effect on the meter pointer. Accurate measurements at one cycle may be made, although there is a slight motion of the pointer.

Analytical studies have indicated that the transient response of the circuit depends to a large extent on the ratio of the interstage coupling capacitor to the output capacitor and moreover that there is an optimum ratio of the phase compensating capacitor to the other two. In altering the amount of suppression therefore, the ratios of the three capacitors are held constant.

The Switching Circuit

Requirements. The basic feature of this measuring system is the switching circuit, consisting of the *A*, *B* and *C* contacts. The requirements for these contacts are: (a) Negligible dc resistance, (b) No contact chatter, (c) Contact potential less than 10 microvolts, and (d) Stability of operation of 20 microseconds. Consideration of these factors led to the selection of mercury contact relays as the basic switching elements. The choice

of relays leads to additional requirements: (e) Substantially equal operate and release times, (f) A make and a break contact, and (g) Substantially zero transfer or bridging time when operated or released.

The first requirement is to avoid correction terms in the test set equations. The second is to make the test set operation conform to the differential equation applying. The third depends upon the fact that meters having dc resistances of the order of 50 ohms are used, and no error deflection due to contact potential effects should be present. The stability requirement results from an objective of studying flux phenomena in intervals as small as 25 microseconds. The added relay requirements will be developed during the switching circuit description. All except the last can be met by individual relays. The last one has not been met by individual relays, but by using contacts of two relays actuated simultaneously, and adjusting their relative timing by winding shunts, the contact switching interval can be reduced to a few microseconds.

Contact Operation. The switching functions required were shown in Figs. 1 (b), (c) and (d) where the interval t_1 can be as low as 25 microseconds or almost as long as T . The A switch is unaffected in duty cycle, but the B and C switches vary from 0 to 100 per cent duty cycle. The operate and release times of relays are affected by their duty cycle and for low or high values become very erratic. To avoid this effect all relays are operated with a 50 per cent duty cycle, and the switching is accomplished by combinations of relay contacts. The epoch of the B relay cycle can be adjusted by manual selection, anywhere between 0 and $T/2$. This is shown schematically in Fig. 8 for the make contacts. The break contacts of course show exactly the reverse behavior. The C make contact is fixed in phase and is set to open just before the A contact closes.

To produce a B cycle for flux rise shown in Fig. 1 (c) where t_1 is less than $T/2$, a parallel connection of the B and C make contacts is used. To pro-

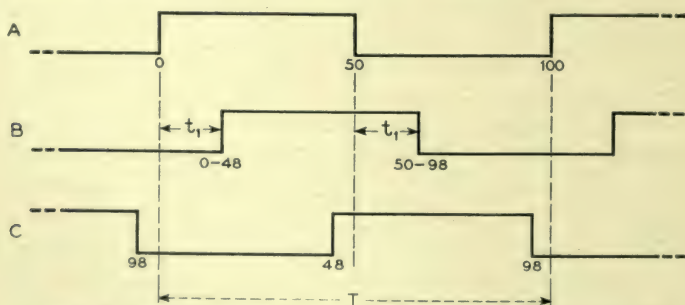


Fig. 8 — The operation of the switching make contacts.

duce a B cycle for flux decay where t_1 is greater than $T/2$, a break contact on the B relay is used in series with the C make contact. Note that in all conditions the meter circuit closure is performed by a B contact and the meter circuit open is performed by the C contact, in fixed phase relationship to the A contact. The addition of the switched meter damping resistor is accomplished by other contacts on the same relays.

A minimum contact circuit² to perform all these switchings is shown in Fig. 9 but was found to be unsatisfactory because of failure to meet requirement (g) above. Only a single transfer switch is necessary to change from flux rise to decay, by transferring the wire marked "x" from the search coil to the damping resistance.

The relays are shown in the released position with the break contacts closed. For the flux rise circuit, contact B can have a momentary open interval when it operates. However, it must not bridge or the damping resistor will momentarily be across the instrument at the initial connection, causing an error. But after contact C operates and then B releases at $(T/2 + t_1)$, an open interval at B will momentarily disconnect the instrument. This would cut out part of the current which must be integrated, causing an error. Thus there are conflicting requirements for flux rise measurements, and transfer contacts of a single relay cannot meet both simultaneously. For flux decay, the B contact cannot bridge because when it releases to connect the instrument, a bridging would momentarily connect the damping resistor across the instrument at the instant of greatest interest, causing an error. The requirements for the C contact are of the same type but less stringent, as the flux is never changing rapidly whenever it is switched.

The foregoing discussion was directed toward establishing the basis for the requirements (f) and (g) above. The former is because of the need for phase reversals in contact operation to perform the necessary

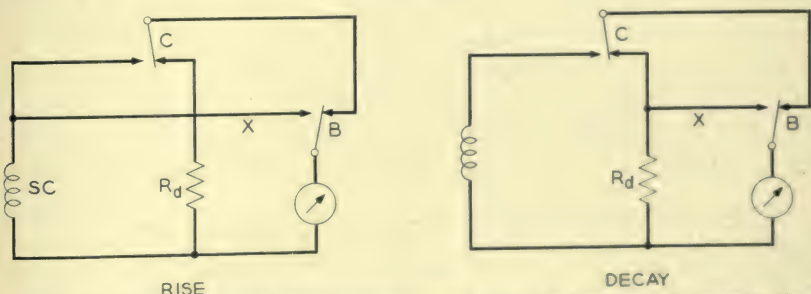


Fig. 9 — Unsatisfactory minimum contact circuit because of transit time requirements.

instrument and damping resistor switching. The latter is to eliminate errors due to momentary circuit configurations not in accordance with the basic equation (3).

The final circuit developed is shown in Fig. 10, which uses two *B* relays, called *B1* and *B2*, a single *C* relay, and compatible timing requirements for the two *B* relays whose windings are connected in series. In all cases *B1* shall be faster than *B2*. The former controls the damping resistor while the latter switches the instrument. A composite circuit is drawn, with the cross marks indicating a closed contact for the function being measured, open otherwise. The change from rise to decay in this circuit requires four switch contacts compared to two for the minimum contact circuit. As before, the relay contacts are all shown in the unoperated position. A preliminary sorting of the *B1* and *B2* relays is required and is based on operate and release time measurements. The faster relay is assigned as *B1*, the other as *B2*. *B2* is further slowed down selectively to match the *A* relay as to operate and release time by the shunts and diode shown in Fig. 10.

B2 can be lined up with the *A* contact both for simultaneous closure of the make contacts when $t_1 = 0$ and for simultaneous closure of the

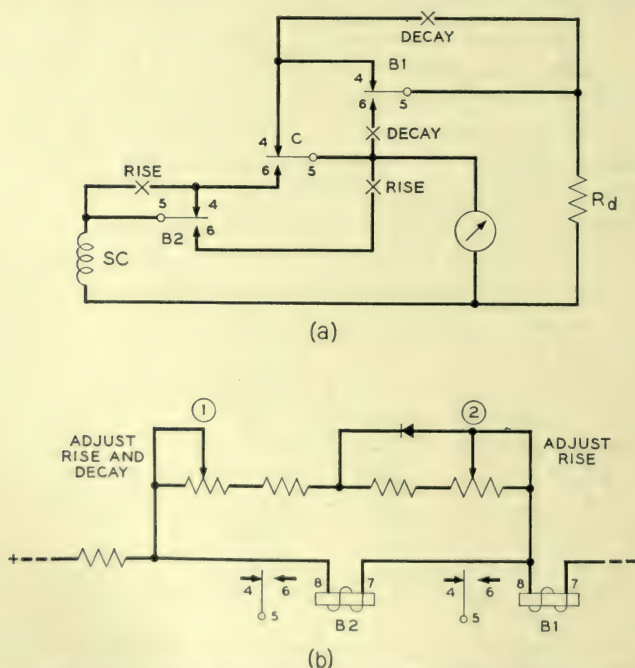


Fig. 10 — Final schematic of basic contact circuit.

break contact of $B1$ with the opening of the make contact of A . The A relay winding has a somewhat similar but fixed set of shunts across it, to slow it down with respect to $B2$ and to equalize operate and release time. The adjustment of the two potentiometers proceeds as follows. The time t_1 is set at 0, flux decay is chosen, and potentiometer 1 is adjusted to the point where the meter just starts to drop from the full reading. Then flux rise is chosen and the potentiometer 2 is adjusted to the point where the meter just starts to rise from zero. A fine check on these settings is to record the meter readings for a few equal small time steps and observe that the successive differences are consistent.

For applying an on and off battery condition as shown in Fig. 9 for the A contact, only one contact on one relay is used. The A contact circuit actually is made of two pairs of relays operated in a push-pull arrangement. The circuit may be changed to connect the four sets of transfer contacts into a lattice configuration to supply battery reversals to the apparatus under test. This is used in testing core materials to eliminate residual effects and for polar relays and similar structures.

The electronic control circuits which apply or stop the relay winding currents will be described later.

Timing Control System

The timing control system consists of a frequency source, a pulse forming circuit, three binary-quinary ring counters in tandem, a time selection circuit, coincidence circuits, memory circuits, and three control circuits, one for each of the A , $B1$ - $B2$ and C relay circuits. A block diagram of the system is shown in Fig. 11.

Oscillator. The frequency source is a bridge stabilized oscillator. This is compared to the Bell System standard frequency for calibration. The accuracy of measurement, both for magnitude and the time scale depends directly upon this oscillator. The magnitude error enters through equation (5) where the cycle time T is exactly the interval for 1000 cycles of the oscillator frequency by virtue of the 1000 discrete steps in the three decades. The time scale is determined by these same steps. Hence the accuracy and stability of the oscillator enters directly; 0.1 per cent can be realized easily. This is as good as necessary as it exceeds the de instrument accuracy.

Pulse Shaper—The pulse shaper is shown in Fig. 12. It consists of an initial cathode coupled amplifier followed by three more direct coupled stages. The square wave output of the third stage is differentiated by a small condenser for voltage spire production. It is a Schmidt³ type circuit designed by I. E. Wood.

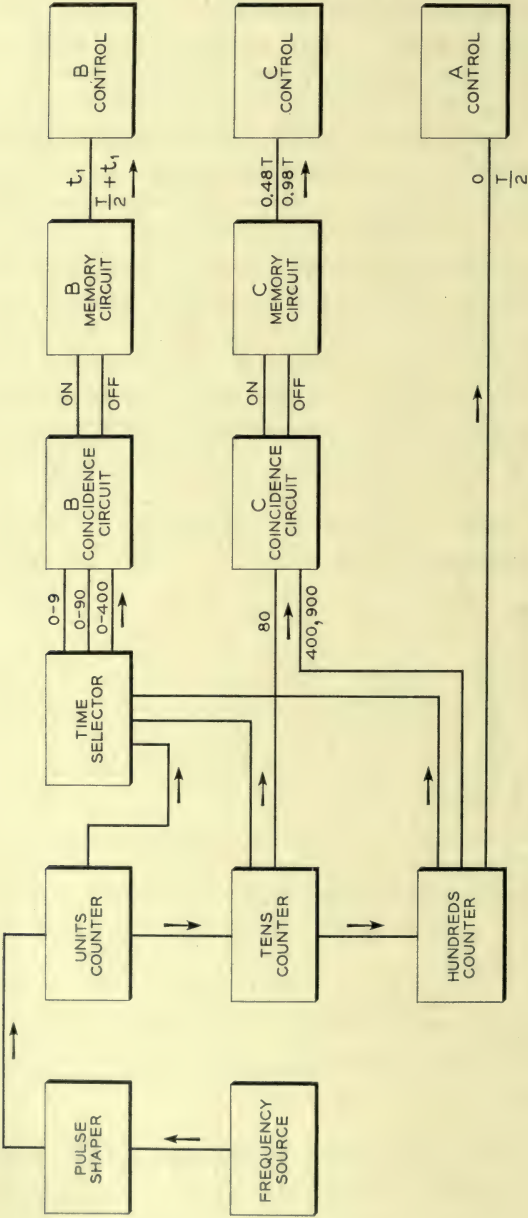


Fig. 11 — Block diagram of timing control system.

The differentiated output appears as a positive voltage impulse across the cathode resistor of a coupling tube, impressed on top of its normal dc bias due to current from the following quinary ring. The coupling circuit has the appearance of a cathode follower, but except during the short intervals of impulse transmission, the plate does not conduct. This lends itself to supplementary use of the rings as counter circuits. An on-off gate can be connected, as shown, to enable or disable the grid bias of this tube, without introducing any false starting or stopping counts.

Quinary-Binary Counters. The unit decade counter is shown in Fig. 13. It is a modification of Weissman's quinary-binary circuit⁴ and was chosen because it provides simple two wire selection for the coincidence circuits. The principal modification consists of applying the shift pulses to the odd cathode lead rather than to a grid multiple, and using a direct coupled positive impulse for shifting.

The shaded tubes are non-conducting at 0 time and can be set in this condition by momentarily opening and then closing the reset key. The plate current from the right half of the zero tube passes through the cathode resistor of the output tube in the pulse shaper described above, and biases both tubes, the latter beyond cutoff. The plate currents from the other four tubes pass through the second cathode resistor having $\frac{1}{4}$ the resistance, developing the same bias.

A positive cathode impulse applied by the pulse shaper shifts the (0) tube from right to left half conduction. In shifting, the negative pulse from the prior non-conducting half of the zero tube cuts off the conducting half of the (1) tube, causing it also to shift. The shift of the (1) tube puts an aiding backward pulse into the (0) tube and a forward pulse into the (2) tube to keep it unchanged. Successive cathode impulses continue the ring stepping, the ring closing on itself after the fifth count. Thus the ability to shift properly depends solely upon the impulse waveform from the pulse shaper, which can be controlled independently of the ring circuit.

At the beginning of the fifth impulse, a square negative impulse is passed by the conducting half of the twin coupling diode to the (5) binary tube, causing it to shift. On the next round, the other half of the diode conducts, restoring the (5) tube. The twin diode behaves as an infinite capacitance in a circuit having zero time constant.

For each of ten successive impulses there is a different configuration of conducting tubes, and two wires, one to the quinary ring and the other to the binary tube, can identify when a selected state is reached by a coincidence circuit noting that a low voltage is present on each wire.

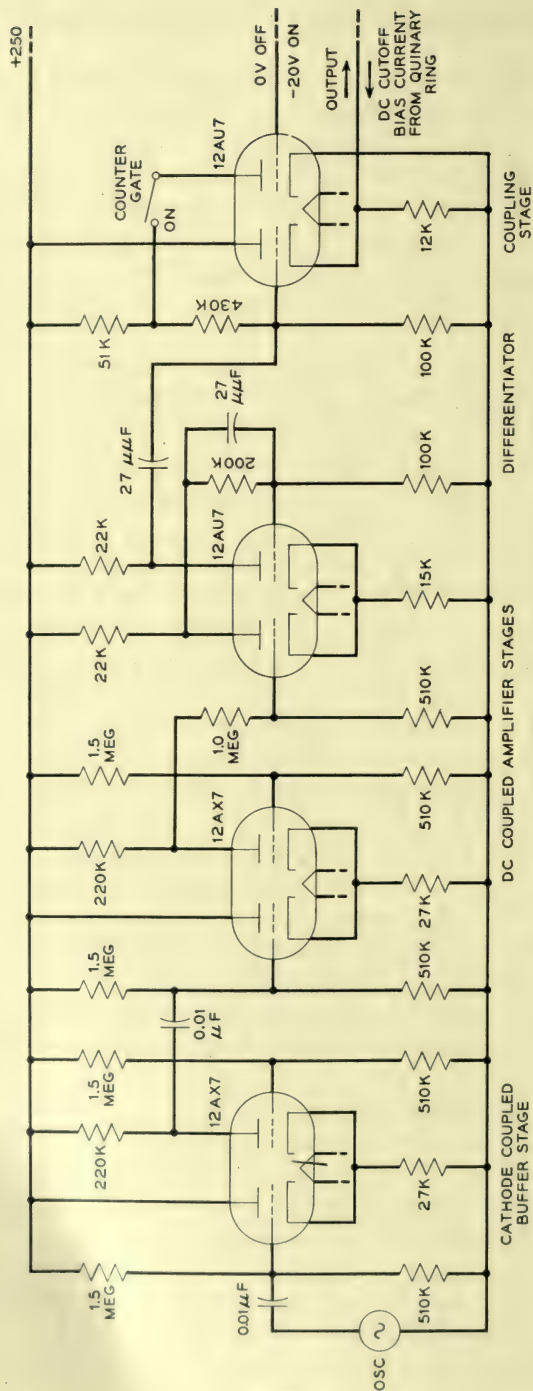


Fig. 12 — Schematic of pulse shaping circuit.

For the other nine configurations, one or the other of the wires will carry nearly the full battery voltage. The leads to the time selector switches are indicated by the arrows marked from 0 to 4, 5R and 5L, inclusive.

When the (5) tube reverts after the tenth impulse, its step plate voltage rise is differentiated by a small capacitor. This drives another pseudo cathode follower to shift the tens counter exactly as has been described for the units counter. For this tens ring, the take off leads to the time selector are identified as 00 to 50.

In a similar manner the hundreds counter leads are identified as 000 to 500.

Thus by exactly six wires, two for each decade, any cycle in 1000 can be selected. Note that for 0 and 500, needed for the *A* relay the (500) tube itself provides complete information. This eliminates the need for a coincidence and memory circuit for the *A* relay drive tube. Also by choosing 480 and 980 for the *C* relay, an abbreviated coincidence circuit can be used, as access to the units counter is not needed.

Neon lamps are provided as indicators for each ring to aid in circuit checking, trouble shooting, and as the counter indicator when used with the gate circuit.

The separation between successive time intervals which can be selected is one/one thousandths of one complete closure and open cycle, being 25 microseconds for a cycle time of 25 milliseconds, 100 microseconds for a cycle time of 100 milliseconds, etc. This is controlled by the discrete states of the counting rings used to generate the switching signals. This relation between the cycle time and the successive available time intervals is not a handicap because if the device is slow and a long cycle time has to be employed, it is slow because the flux buildup or decay is slow and hence closely spaced measurements are superfluous.

The maximum speed is 40 on-off cycles a second obtained with a 40-ke oscillator. The lowest speed is limited only by the ability of the vacuum tube filter to suppress instrument pointer vibration.

The counting ring system with its discrete steps precludes an automatically recording device as would be possible with a motor driven commutator and a gear driven take-off brush. However, the elimination of brush troubles is considered to be worth this sacrifice.

Time Selector. The time selection is controlled by two two-gang decade switches for the units and tens selections and one five position switch for the hundreds selection. The schematic is shown in Fig. 14. The dial positions are marked directly in time for a 10-ke oscillator, that is the units selector indicates one-tenth milliseconds, the tens milliseconds, and

the hundreds tens of milliseconds. For other oscillator frequencies the times indicated are scaled in proportion. Thus for a 40-kc oscillator, the indicated times are divided by 4 and the fine steps become 25 microseconds.

The start of the *A* cycle is marked by L500 and the stop by R500, the 500 tube itself serving as the memory circuit, and no coincidence circuit is needed.

The *C* cycle is wired permanently with a phase to close just before 500 and open just before 0. By choosing 480 and 980, access to the units decade is unnecessary and only four coincidence circuits are provided.

The start of the *B* cycle is marked by six coincidences, the end by five of these and the alternate connection to the 500 tube. The hundreds switch is simplified because the *B* relay always operates between 0 and 500 and releases between 500 and 1000. Therefore, the start and stop coincidence circuits can be wired permanently to the 500 tube and only

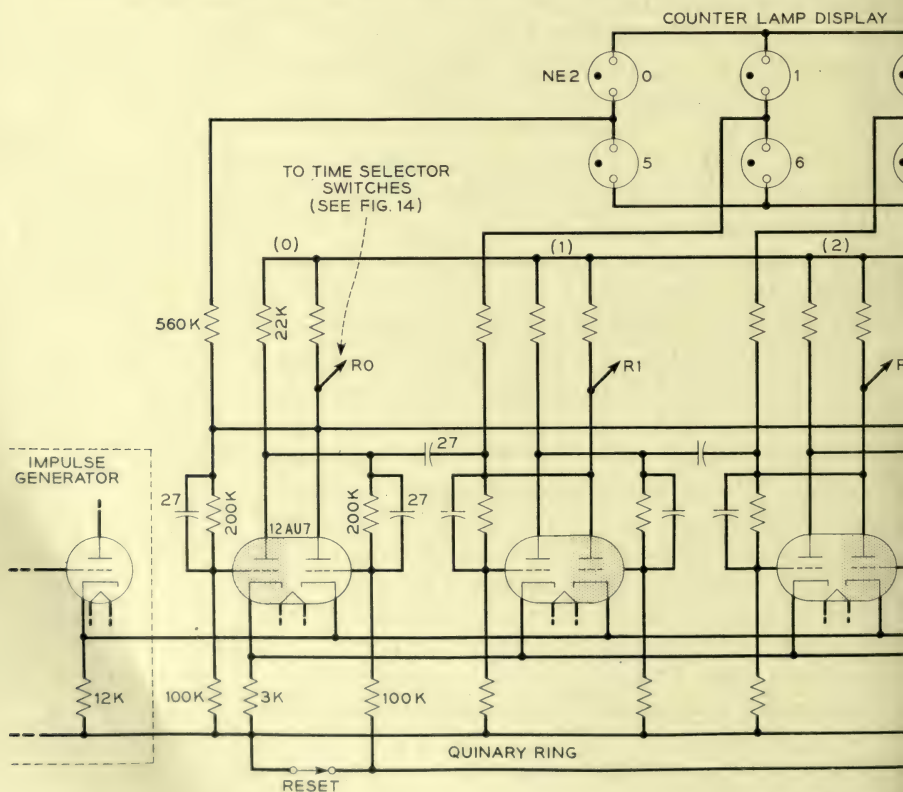
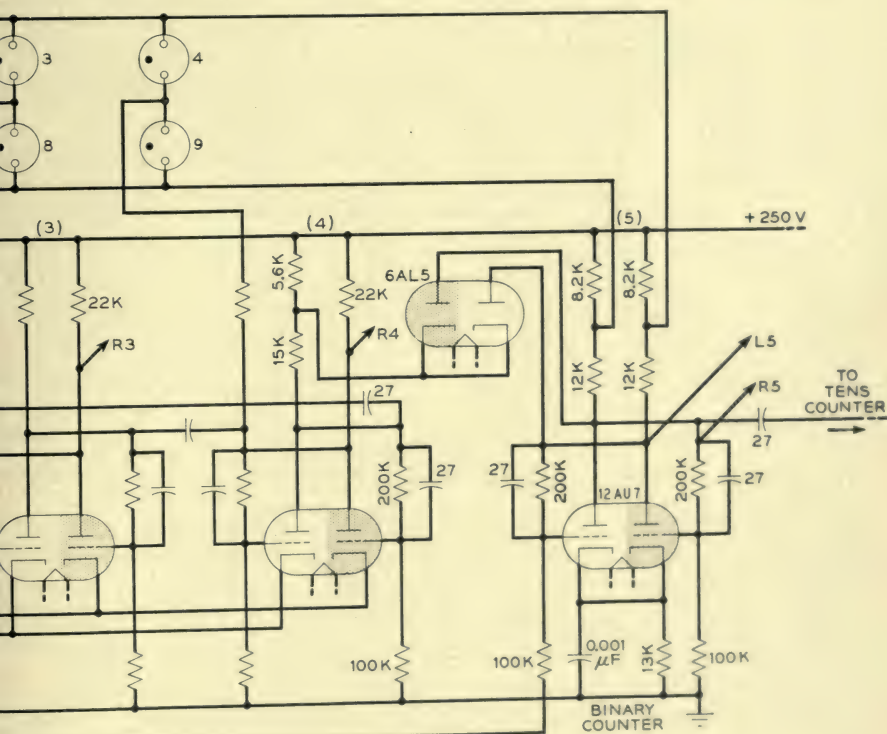


Fig. 13 — Inte

five switched wires are necessary. The change from flux rise to flux decay by another switch as shown in Fig. 10, converts the contact switching to provide the proper instrument closing time, with the time dials now reading time from the beginning of the open of the *A* circuit.

Coincidence Circuits. Each coincidence circuit consists of triodes, one for each marked wire. The circuit for the *B* cycle is shown in Fig. 15. Each selector switch wire connects to a grid, but all the cathodes of a set of coincidence tubes use a common cathode resistor. Except for the instant chosen, at least one coincidence tube has a high potential on its grid and the plate current through that tube holds the cathode resistor at a high potential. Only at the chosen time are all plates of the counter tubes conducting and therefore at a low potential. This abruptly drops the cathode potential of the coincidence circuit for one cycle of the oscillator, recurring every 1000 cycles. 500 cycles after each start pulse a similar stop



er schematic.

pulse occurs in a duplicate set of coincidence tubes. These alternate pulses control the state of the memory circuit.

The *C* cycle coincidence circuits are similar except only four triodes are used for the start, and four for the stop pulses.

Memory and Relay Control Circuits. The memory and control circuits are shown in Fig. 16. The function of the bi-stable memory circuits is to accept the alternate start and stop pulses from the coincidence circuits, switch to the state representing the imposed condition, and hold it until the next successive imposed condition.

The function of the relay control circuits is to operate or release the relays under control of the memory circuits. The relay control circuits are direct coupled to the memory circuits and one or the other plate is cut-off by the grid bias condition imposed. The relays are in the plate circuits of the control tubes and either have full or no current applied. The operate and release times of the relays add a delay in the contact

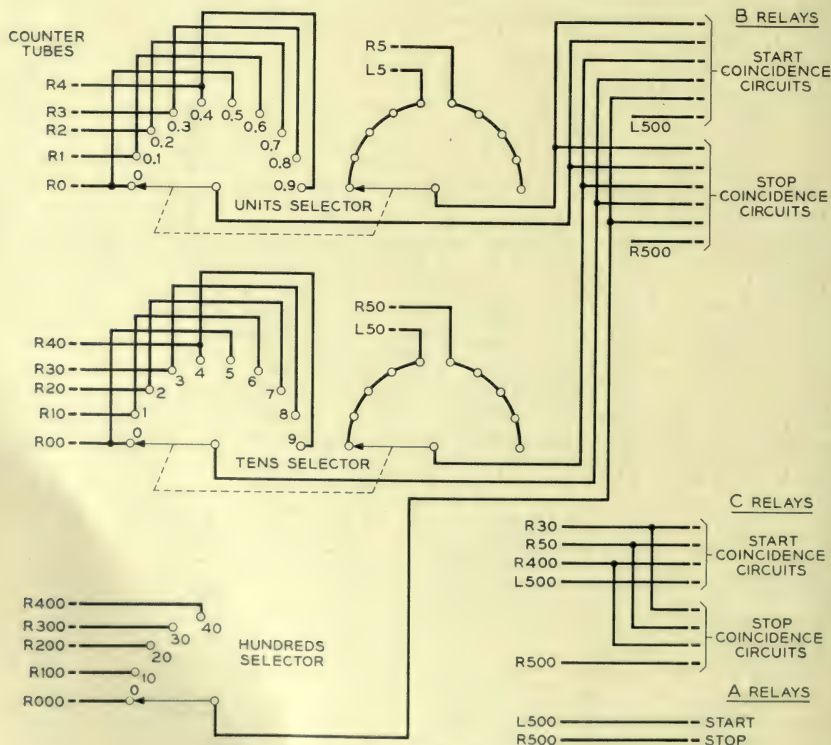


Fig. 14 — Time selector schematic.

actuation compared to the state of the memory tube. By making these small and equal, through selection of the relay design, the constant time lags do not alter the relative preset switching instants.

The relays are special mercury contact, single transfer, Western Electric 291-type. They have 38 gauge 1500 ohms, 14,600 turns windings and are adjusted to operate with 3.5 milliamperes and release with 2.0 milliamperes. The 5687 tubes switch from 0 current to 10 milliamperes through the relays. The release and operate times of the relays are about 1.3 milliseconds but as described earlier, are selectively adjusted to 1.5 milliseconds where necessary, by shunts such as those shown in Fig. 10.

The bi-stable memory circuits are switched at the grids through the twin diodes coupling the coincidence and memory circuits. While waiting for a shift pulse, the grid potential of the memory tube is lower than the cathode potential of the coincidence circuit, and the diode circuit therefore does not conduct. This is necessary because small voltage variations occur in the coincidence circuit cathode potential as the several tubes follow the decade counters. These small variations would shift the memory circuit falsely if directly connected. When the shift coincidence occurs, the cathode voltage drop is to a potential lower than that of the memory circuit grid and the conduction of the diode connects the two circuits. This negative impulse shifts the memory circuit. This drives the start grid to a new potential lower than that of the coincidence tubes cathodes, again cutting off the connection through the diode. When the coincidence cathode rises at the end of the shift pulse, the diode is even further cut off. The stop diode behaves in an exactly similar manner due to the symmetry of the circuits. The diodes thus afford a means of only momentarily making connections at the required instants, at all other times isolating the memory circuit from the remainder of the counting system.

Summary of Dynamic Fluxmeter Description

Up to this point, the circuit description has been concerned only with the dynamic fluxmeter. This circuit was designed and built first, and put into operation for dynamic current and flux studies of which one will be described later. About fifty standard type tubes in all are used, with good segregation of the circuit functions. This aids in localizing circuit troubles. Particularly useful is the neon indicator lamp display, also employed when the decade system, with the gate circuit, is adapted as a counter or time interval circuit.

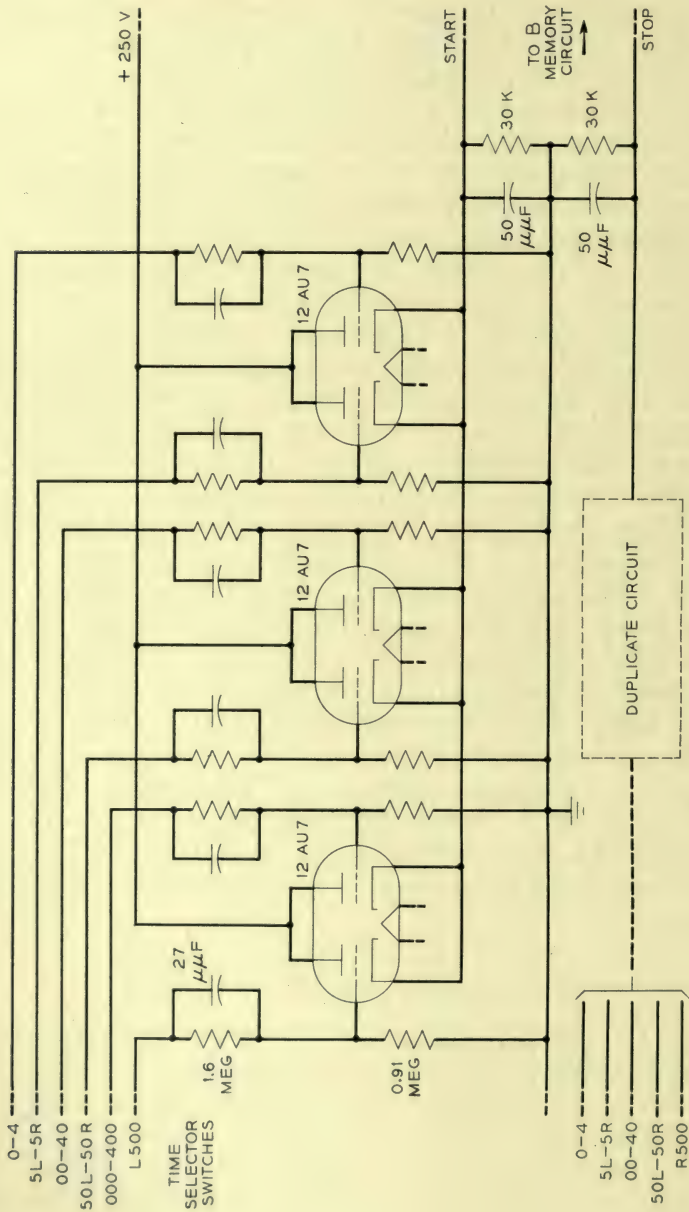
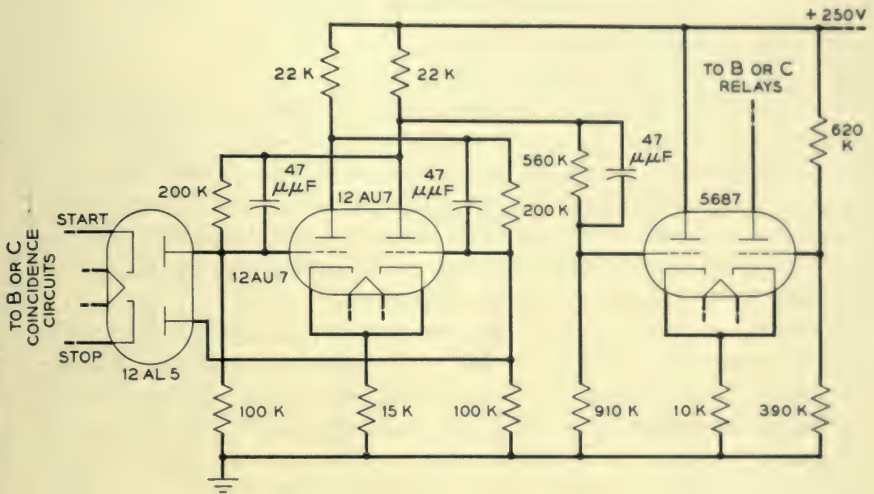
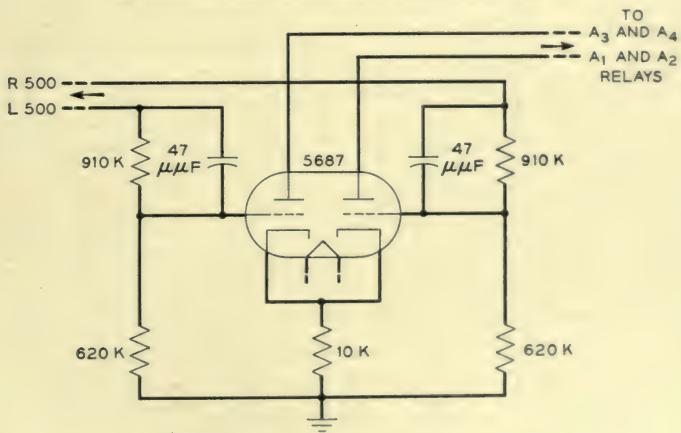


Fig. 15 — B-Cycle coincidence circuits.



B OR C MEMORY AND RELAY CONTROL CIRCUIT



A RELAY CONTROL CIRCUIT

Fig. 16 — Memory and relay control circuits.

Part II — Displacement and Velocity Measuring System

*The Optical System*¹

A block diagram of the displacement and velocity measuring system was shown as Fig. 4. As part of this figure, a geometric schematic of the optical part of the system is shown. Most of the features of the optical system are due to the advice and assistance of associates in these laboratories who were connected with the sound motion picture development. The fundamental design is similar to the film reproducing system with certain necessary changes in dimensions.

Referring to Fig. 4, the elements are the lamp, condensing lens, a slit, objective lenses, a vane on the part whose motion is to be studied, and a photocell. The lamp and the condensing lens are the same as are used in the film reproducer. The slit is the same except that its width has been increased to 0.005" and a feature added permitting its length to be adjusted from zero to about 0.1". The objective lenses are inexpensive Bausch and Lomb achromatic lenses. The second objective lens is interchangeable for different focal lengths, and may be moved back and forth in the supporting tube for precisely focusing the image of the slit on the moving vane.

The three lenses and slit are mounted in a tube fastened to the lamp housing which in turn is supported by adjustable supports mounted on a vertical stand. To the lamp housing support is also fastened the photocell container with its first amplifier tube.

A permalloy shield surrounds the photocell except for an opening for access of light which passes by the moving vane. The shield is provided to prevent the changing stray magnetic field from the nearby electromagnet under test from affecting the photocell current. This system can be moved in any of three directions for lining up with the vanes of the device being tested. The relay or structure being measured is not fastened to the optical system support, but is secured to another appropriate stand resting on the same laboratory bench.

This system may seem more complicated than necessary at first sight, but its advantage is that it provides a rectangular beam of light of constant width and adjustable length at the point being measured. The necessity of numerous light shields and screens is done away with, although it is advisable to throw a black cloth over the whole apparatus when once adjusted to keep out the overhead illumination.

The reason for this last is of interest. The fluxmeter timing is controlled by a stable oscillator. The 60-cycle power for the overhead lights is also

well controlled. The two systems therefore operate in substantially a synchronized fashion when convenient cycle times are chosen, such as 10 cycles per second. The overhead fluorescent lamps fluctuate in light intensity at 120 cycles per second, so that the contacts in the fluxmeter operate in synchronism with the fluctuations in light intensity. Any measurements made with overhead illumination getting into the photocell will therefore have a superimposed 120-cycle ripple. For the same reason extreme care has to be taken to avoid any 60-cycle pickup in the apparatus. The lamp and the heaters on the first amplifier tubes are supplied with well filtered dc and cannot be operated by ac. Noise not in synchronism with the contacts is of little importance, as the averaging of the meter cancels it. A comment on the linearity of the photocell should be made at this point. This system operates on a variable width, rather than a variable density basis. Consequently the linearity of the system depends upon the uniformity of emission of the photocell surface. This can be verified by moving a vane with a micrometer and recording the output of the amplifier with a precision dc voltmeter. If the relationship is not linear other photocells can be substituted until sufficient linearity is achieved.

Amplifier System

General. Fig. 4 showed the three dc amplifiers in block diagram form. These amplifiers have each been designed to operate with full internal gain from dc to 10,000 cycles. The external transfer characteristic of each is controlled by its input and feedback networks. The design of these networks provides ideal transfer characteristics from dc up to 10,000 cycles. At 10 kc, the frequency response deviates by less than 3 db from the ideal. This same frequency corresponds to a time constant of 16 microseconds. On a small signal basis, the fidelity of measurement therefore will extend to events occurring in times of the order of 16 microseconds.

A more serious limitation to the accuracy is the finite plate voltage available for the output amplifier which results in amplifier over loading on large peaks. As will be shown, this can delay the response of the meter to sudden velocity discontinuities.

The amplifiers have been designed to operate from two voltage supplies, plus and minus 250 volts. The -250 volts is a series regulated three stage circuit with an output impedance of less than 0.8 ohm at all frequencies. It also serves as the reference voltage for a three stage shunt regulated +250-volt supply. Keeping the magnitudes equal minimizes errors due to power supply voltage variations.

The photocell amplifier converts the current from the high impedance photocell into a proportional voltage, having an internal impedance of 10 ohms.

The following amplifier-differentiator has an internal gain of 80 db from dc to 10 kc. When used as a differentiator the rising external gain characteristic reaches 67 db at 10 kc.

The output amplifier also has an internal gain of 80 db from dc to 10 kc. The feedback network includes an equalizer to produce current, and hence flux, in the air core output transformer proportional to input voltage from dc to 10 kc. The inductance of the transformer, with constant applied ac voltage, would cause a 6 db per octave decrease in current above the frequency where its Q is unity. This is counteracted by designing the external amplifier gain to increase at 6 db per octave, starting at the same frequency. The output amplifier thus has the same characteristics as a differentiator at high frequencies and its external gain is 64 db at 10 kc.

The overall gain of the system, then, is 131 db at 10 kc. Stability has been obtained, even with this much overall amplification without resort to compartmentation. Each tube stage utilizes shielded turret construction, exposed interstage leads are very short and only the input grid leads are shielded because they connect to controls on the front panel. A box shield surrounds the gain selector and the displacement-velocity switch of the middle amplifier, the lowest level point on the main panel.

This system uses an electronic differentiator. Ordinarily in analogue computers, these are avoided because of the high-frequency noise introduced as a result of the attendant large amplification. In the present system, the averaging of the succession of impulses by the dc instrument minimizes this effect. With each cycle a discontinuous section of the noise is sampled, containing a dc component, but as these are random in sign, their average gives rise to no error.

The system is dc coupled up to the output transformer. Slow drifts, due to grid currents or other reasons, do not result in instrument current errors because of the differentiating action of the transformer. No actual dc source appears in the meter circuit itself. The only concern here is to maintain the amplifiers somewhere near their best operating point. Actually the main source of dc drift is the temperature coefficient of the photocell and the voltage stability of the lamp power supply.

Calibration. The calibration of the system starts with a static measurement of the total displacement of the reciprocating motion to be studied. This can be done using thickness gauges or a tool maker's microscope. Then the device is brought into alignment with the light beam and cycled

by the *A* contact. The cycle time is chosen for complete operation and the time selector is set for a time near the end of the operate interval. At this time the tested device is known to be at its maximum displacement, and the instrument reading, using the displacement connection, corresponds. A convenient instrument indication is obtained by adjustment of the amplifier gain controls. For instance if the displacement were 0.040", an instrument reading of 200 microamperes could be used. At any other time, when the parts are in relative motion, the instantaneous displacement is read directly from the instrument using the same scale conversion factor.

The calibration for velocity measurements depends upon the displacement calibration and relationship between the input differentiating capacitor and the resistor it replaces. Once a displacement gain setting has been chosen, it must not be altered during the associated velocity measurements except by calibrated steps. The differentiating capacitor has been chosen so that an instrument displacement deflection, corresponding to a given number of thousandths of an inch, represents the same number in inches per second. For the example given above, a 200-microampere instrument reading would represent an instantaneous velocity of 40 inches per second, with a linear calibration for intermediate readings.

A plot of measured displacement and velocity for a fast wire spring relay shown in Fig. 21, will be described when system errors are considered.

Photocell and Impedance Transformer Amplifier. A schematic of the photocell and dc impedance transformer is shown in Fig. 17. The high vacuum photocell has an impedance of thousands of megohms and acts essentially as a constant current device, the current depending upon the instantaneous illumination. This current is connected to the grid of a series stabilized⁵ twin triode amplifier tube, to which grid also is connected the current through the feedback resistor and a balancing current from the -250-volt power supply. The zero adjustment is made under quiescent conditions for the desired dc output voltage, there being, of course, essentially no dc current in the grid.

The Western Electric low grid current 420A input tube is stabilized both for heater voltage and plate potential. It provides a voltage amplification of exactly $\mu/2$ or 35, by virtue of the upper tube having an impedance exactly equal to that of the plate of the lower tube. This input tube is mounted immediately behind the photocell in the same shield.

The output from the first amplifier tube is connected through a double shielded cable to the output tube mounted on the main chassis. The 5687 twin triode output tube uses both halves as cathode followers, one

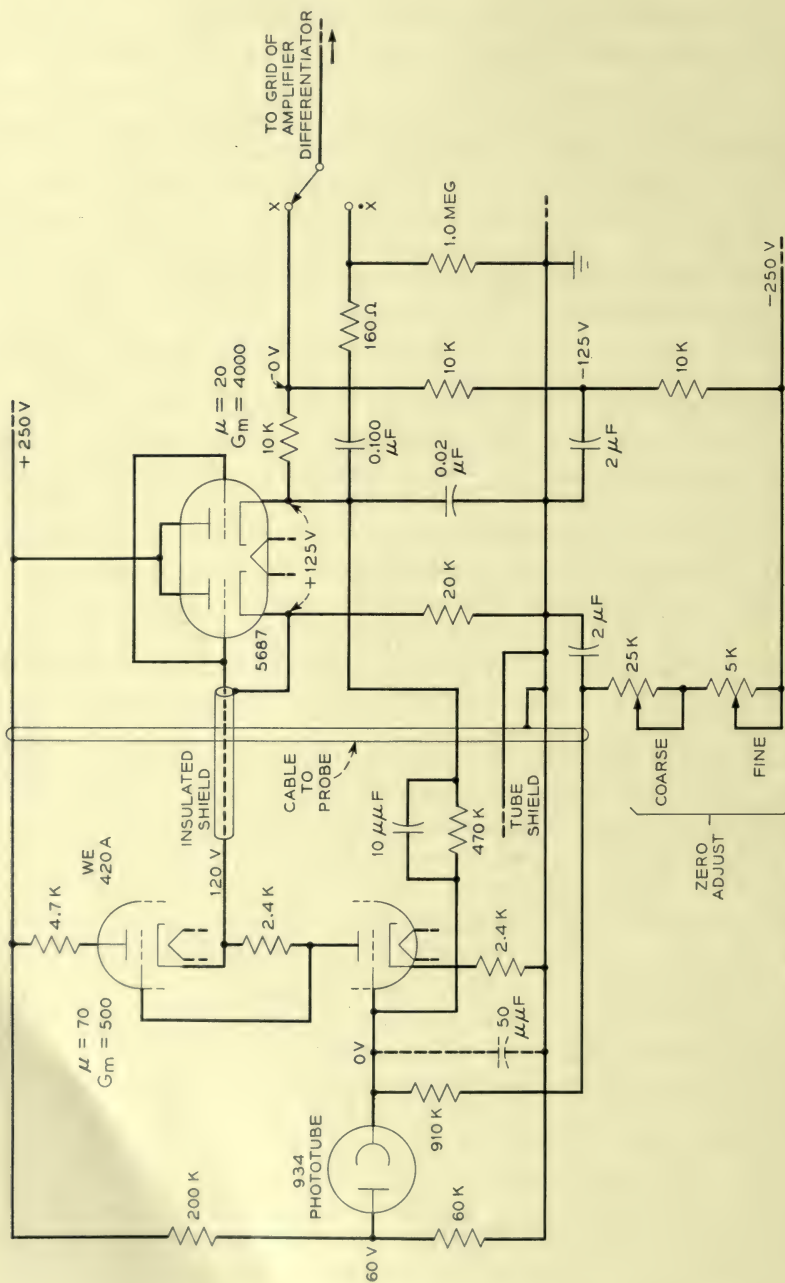


Fig. 17 — Schematic of photocell and DC impedance transformer.

to provide the output voltage and the other as a driver for the insulated inner shield to reduce the effective capacity to ground of the interstage cable connection. The cathode potential of the output tube is operated at exactly +125 volts, set by means of the zero adjustment previously described.

The input networks of the succeeding amplifier-differentiator are shown as part of this circuit as they provide the path for the output tube current and enter directly into the frequency response and loop gain cutoff design of this feedback amplifier. The dc grid voltage of the succeeding amplifier is 0, and this potential is found one-third the way down the cathode resistor which connects to the -250 volts. This point then provides the input to the following grid for displacement measurements and the input resistance for gain computations of the following amplifier is only this one-third part of the total cathode resistance or 10,000 ohms.

For velocity measurements, the succeeding grid is switched to the polystyrene differentiating capacitor whose grid side also is kept in readiness at 0 voltage by a grounded 1-megohm resistor.

For a change in current from the photocell due to a change in light, the feedback acts to supply an equal but opposing current from the output to keep the input grid nearly at its virtual ground potential. Thus the output voltage change E_s is given to a good approximation by the simple relation

$$E_s = \frac{I_s R_f \mu \beta}{1 - \mu \beta} \approx -I_s R_f, \quad (16)$$

where I_s = change in photocell current, R_f = feedback resistance, and $\mu \beta$ = loop gain.

A word about the differentiator connection is in order here. For this use, the cathode load approaches 160 ohms at high frequencies because the input grid of the next amplifier is a virtual ground point, and the load becomes equal to the phase shift controlling resistor in series with the 0.1 mf capacitor, the value of which will be discussed later. The impedance of the cathode follower without feedback is 250 ohms but this does not mean that it can be connected to a 250 ohm load and operated at the normal power output rating. The basic limitation for any tube is the allowable plate current change, regardless of the external load, consistent with never drawing grid current or being cutoff. For instance, in a cathode follower, if a load equal to $1/G_m$ is used, the small signal output voltage is only half of the applied grid voltage, a 6 db loss. Lower load resistances result in corresponding greater losses. In the present circuit, about an 8 db reduction in loop gain occurs at 10 kc

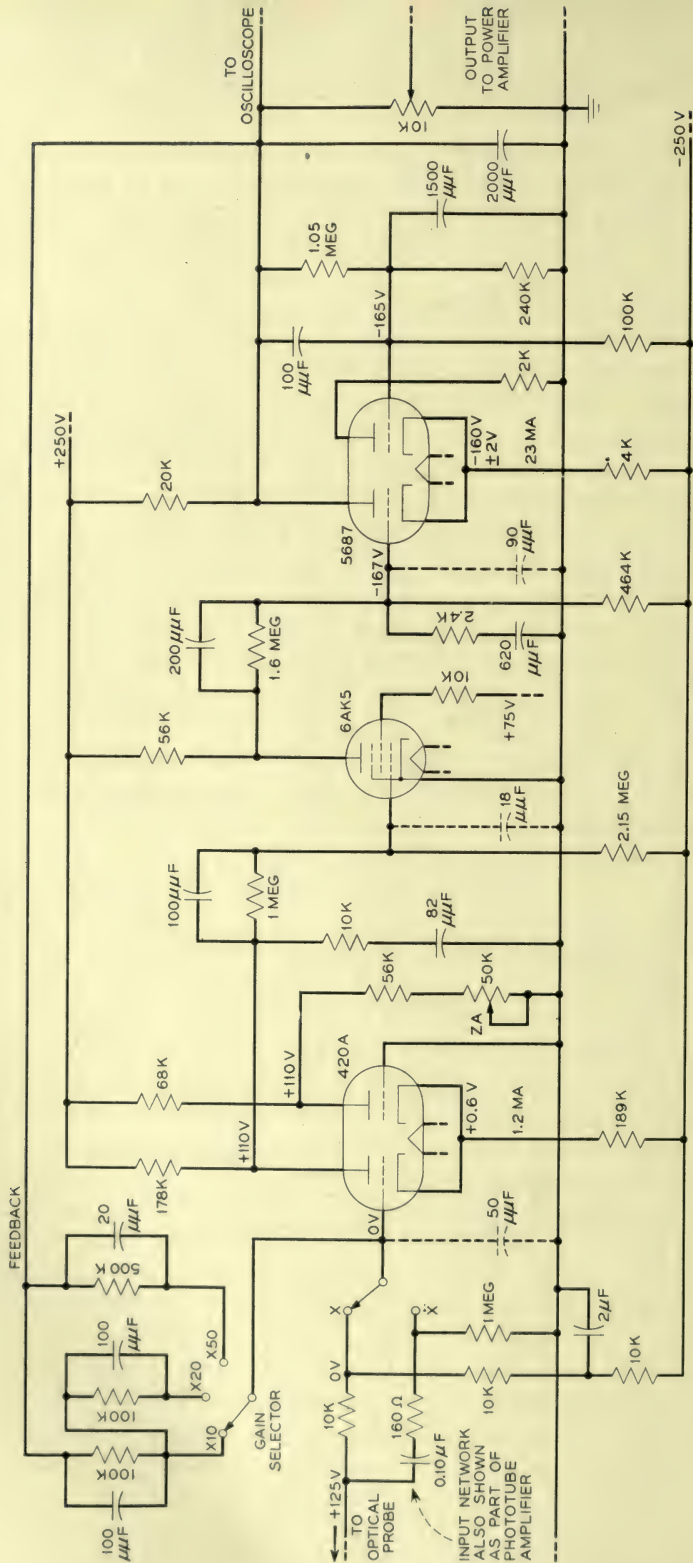


Fig. 18 — Amplifier-differentiator schematic.

leaving a net gain of about 22 db, including the 3 db effect of the input grid capacity.

Of equal concern is the reduced power handling capacity, but fortunately the energy in the frequency spectrum of the input signal diminishes rapidly with increasing frequency. A measure of this can be arrived at by the following analysis. The output voltage change for a motion of 0.040" is about 0.1 volt. This total motion ordinarily never takes place in less than about 0.001 second, or a maximum rate of change of 100 volts per second. Now for a capacitive circuit

$$i = C \frac{dE}{dt} = 0.1 \times 10^{-6} \times 100 = 10 \text{ microamperes,} \quad (17)$$

which is extremely small compared to the 12.5 milliamperes quiescent cathode current. Thus in choosing a vacuum tube to provide the desired amplifier output impedance characteristics, all other considerations have been cared for.

Amplifier-Differentiator. A schematic of the amplifier-differentiator is shown as Fig. 18. It has three stages, is direct coupled, and the quiescent output voltage is zero, set by the zero adjustment which controls the plate voltage of the inner tube of the input stage. The first stage is a conventional parallel stabilized circuit. It uses a high common cathode resistor to return the twin plate currents to the -250-volt supply. The first stage output is fractionated by an L pad to provide the proper dc bias for the second pentode stage. The output of the second stage is likewise fractionated for the dc bias of the final stage.

The third stage is a twin triode designed with an inside positive feedback loop. The two tubes have a common cathode resistor. After arbitrarily choosing one of the three grid resistors of the inside tube, the other two can be determined so that the proper quiescent bias and voltage changes result to make the plate current of the inner tube complement the plate current of the output triode itself. This maintains the total cathode current constant and establishes the cathodes as a virtual zero impedance to ground point. This realizes full gain from the output stage without another power supply. It also protects the triode from destruction if an accidental ground connection is made to the output.

The feedback circuit permits a choice of three values of resistances. The external amplification can be set at 10, 20, or 50. The maximum value of feedback resistance which can be used is determined by the value of the differentiating capacitor. It is set by the requirement that at least 10 db of loop gain remain at 10 kc.

It is now appropriate to discuss the scale factor for the differentiator

and to arrive at the value of capacitance to select. First we wish to measure reciprocating motions of the order of 0.050" and actuating times of the order of 0.005 second. However, the motion takes place in the order of half the actuating time. Hence the average velocity will be of the order of 20 inches per second. An average numerical ratio of velocity to displacement then is 20 divided by 0.05, or 400. Remembering this is average and some part of the velocity will be higher the nearest decimal factor is 1000.

In this system, as a device being tested is a part of it, the instrument time scale necessarily is identical with real time, that is it cannot be "slowed down" for more accurate time settings, as can analogue computers.

Now once a convenient gain setting for displacement measurements has been determined, if a capacitor were substituted for the amplifier input resistor, related by the equation $RC = 1$ second, C would be 100 *mf*, as $R = 10,000$ ohms. This would be a substitution where the scale reading for a given number of mil-inches would represent the same number of mil-inches per second. Fortunately, however, we wish rather to represent a much higher velocity of inches per second. This results if the capacitor is reduced by the same factor of 1000. The desired capacitance thus has a value of 0.10 *mf*. The type used is a stable, low soak polystyrene dielectric one having $\frac{1}{2}$ per cent accuracy. In this amplifier the feedback and input resistors also must be of $\frac{1}{2}$ per cent accuracy as upon these depend the calibration. The 10,000 ohms input resistor necessarily is a non-inductive wire wound type with a 10 watt rating, for stable performance with the actual 1.5 watts. The feedback resistors are of the deposited carbon type.

The remainder of the design now follows. To limit the differentiation action to a 10-kc range, a resistor in series with the 0.1-*mf* capacitor of 160 ohms is used. A transfer gain of 70 db, which is 10 db less than the internal gain of 80 db, is a voltage ratio of 3,160. Multiplying this by the input resistor value of 160 ohms yields 500,000 ohms as the maximum feedback resistance. Then as an amplifier with an input resistor of 10,000 ohms, the gain is 50. Resistances for lesser gains of 20 and 10 are scaled in proportion.

The linear output voltage range for this amplifier is ± 40 volts. This is much more than needed. For displacement measurements the maximum swing is about 5 volts. In the discussion, we found a maximum capacitor current of about 10 microamperes which also results in an output swing of 5 volts, during velocity measurements.

Output Amplifier. The schematic of the output amplifier is shown in Fig. 19. This amplifier also has three stages, the first two of which are

like those of the amplifier just described. Similar considerations for the overall design were used here and the loop gain cutoff control uses the same principles, modified to take account of the inductor load and its constant current feedback equalization.

The output tube is a 6L6 with an 874 voltage regulator tube in the cathode return to the -250 volts. The 6L6 is operated at 40 milliamperes, which is within the current rating of the 874. This stage was designed for the widest possible output voltage swing and for plate current cutoff independent of plate voltage, both characteristic of pentodes.

The transformer has a 900-cycle self resonance with its equivalent shunt capacitance. At higher frequencies it behaves as a capacitor as far as the tube and loop gain are concerned. Its Q is unity at 56 cycles.

The function of the output amplifier is to convert the input voltage signal into proportional current in the inductor. Constant voltage gain would not do this. For frequencies above 56 cycles, the current would drop at a 6 db per octave rate with increasing frequency, because of the reactance.

Note that if the voltage across the inductor rises with increasing frequency, then the current through the winding will be proportional to a constant input voltage regardless of the equivalent shunt capacitor. The capacitor will draw greatly increased current but as long as it is uniformly distributed, it will not affect the winding current itself. If series output feedback were used, the two currents could not be separated and a more complex equalizer would be necessary to provide the necessary frequency response. The choice of the value of inductance for the output transformer will be discussed after errors in the system have been considered.

The rising gain is produced by shaping the feedback, starting an insertion gain rise at 56 cycles with the 0.057-mf capacitor, and continuing to 10 kc. At this frequency the rising characteristic is discontinued by the 280-ohm series resistor, and the capacitors shunting the feedback resistors. Over the rising characteristic range the equalization introduces 90° adverse loop phase shift and above 900 cycles the output transformer adds another 90° phase shift. Thus from 900 cycles to 10 kc the amplifier just skirts being Nyquist stable.⁶

The internal gain of this amplifier is 80 db. At dc the external gain is 21 db. It rises to 67 db at 10 kc because of the feedback equalization.

Output Amplifier Square Wave Response

The square wave response of the output amplifier is shown in Fig. 20. Two effects are evident (1) a finite time for the major change to occur,

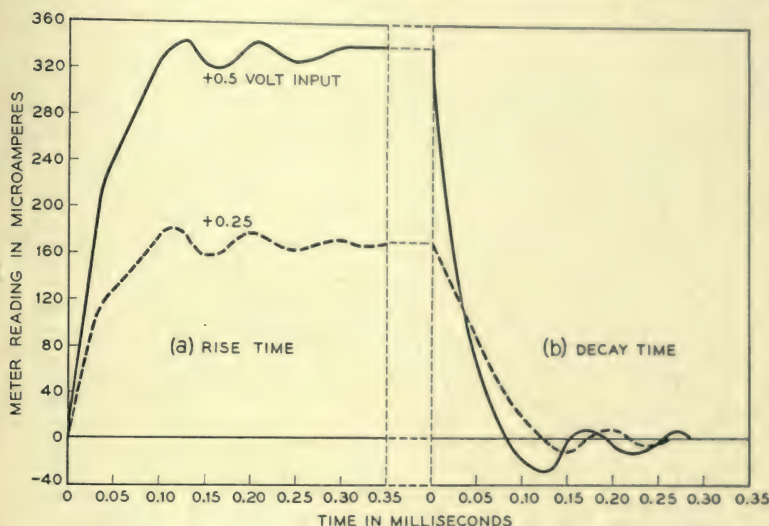


Fig. 20 — Output amplifier square wave response.

followed by (2) an oscillation caused by the frequency response of the amplifier deviating from the ideal for frequencies above 10 kc. The former effect is caused by the finite power supply voltages available in the final stage of the output amplifier.

The current in the primary of the output transformer represents the function being measured. If a step discontinuity occurs, as during velocity measurements, then the current suddenly has to be changed to a different value. The rate of change of an increase in current which can be developed is proportional to the power supply voltage, and hence is finite. For a decrease, the distributed capacitance of the winding and the shunting resistors delay the current decay. These result in a delayed transition from one condition to the other, occurring, from Fig. 20, in about 0.1 millisecond.

A discussion of the operating conditions of the system leads to a method of evaluating these forms of error and determining what limitations are imposed upon the use of the measured data. It will be shown that only immediately following velocity discontinuities are the data not usable. Displacement data are never in difficulty from an overload standpoint and fortunately we ordinarily are not too interested in velocities after impacts.

Discontinuity Errors Due to Overloading

The arbitrary gain setting for the output amplifier is chosen to provide about the full scale of 200 microamperes to represent the full displace-

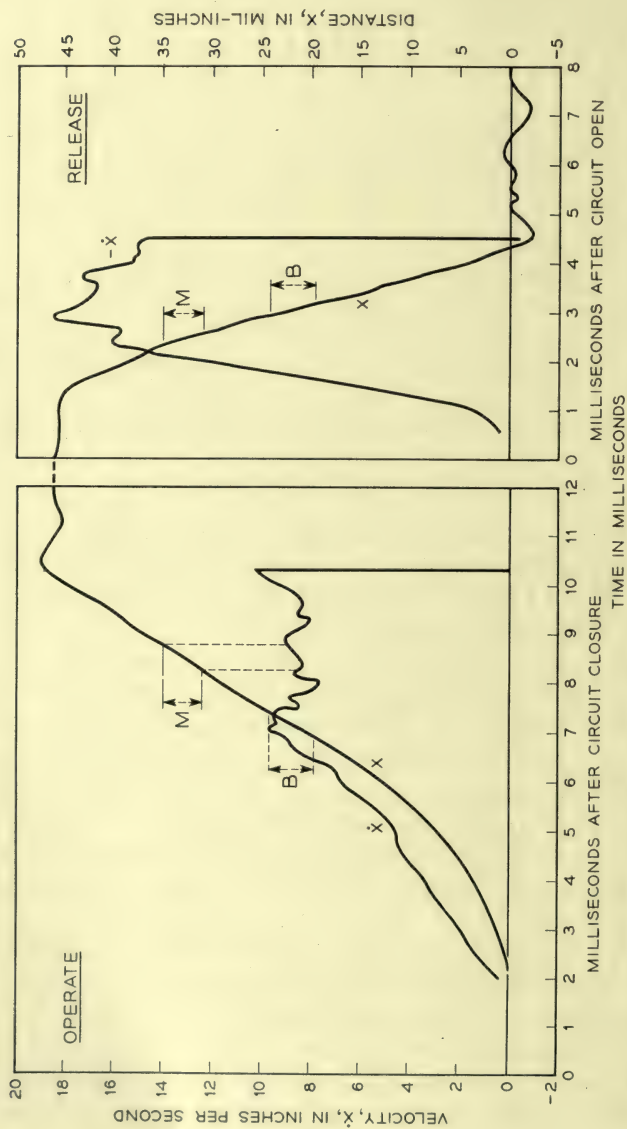


Fig. 21 — Displacement and velocity measurements of the AF-type relay. M — on operate, make contacts closed in this region, and on release, make contacts open in this region. B — on operate, break contacts open in this region, and on release, break contacts close in this region.

ment. The output transformer obeys equation (8), and its characteristics are tabulated on Figure 19. From this equation and a cycle time of 0.1 second we find that the change in plate current for a 0.705 henry mutual inductance amounts to 2.83 ma. The dc plate voltage change therefore is 11.32 volts and the corresponding input dc voltage is 1.0. These values are extremely small compared to the voltage swings available, but nevertheless the fact that the instantaneous voltage swings are finite imposes a limit to measurement of step discontinuities.

One direction of transformer current change is where we wish to decrease the current suddenly. The signal blocks the output tube and the transformer current decays through the equivalent of a 50,000-ohm shunt. The primary inductance is 11.5 henries and the time constant therefore is 0.23 millisecond. The quiescent current is 40 microamperes, so a drop of 2.8 microamperes, assuming the 6L6 blocks completely, would require at least 0.02 millisecond. This means that if there is a measurement discontinuity, the best the circuit can do is to respond as the first part of an exponential over at least this time interval, rather than abruptly. This is shown in Fig. 20, where the indicated time actually approaches 0.1 millisecond. This extension evidently is contributed to by the distributed capacitance of the winding.

The current rise case is about as favorable. The current rise rate is limited by Lenz's law to a maximum value of

$$\frac{di}{dt} = \frac{E}{L} = \frac{265V}{11.5H} = 23 \text{ ma/ms.} \quad (18)$$

A change of 2.83 microamperes therefore requires at least 0.12 millisecond. This is also shown in Fig. 20.

The foregoing discussion has been directed toward establishing that the initial delays of Fig. 20 are entirely overload effects and not a frequency response effect. Thus if required rates of change during a test do not exceed the available rate, then no error of this type will occur.

Fig. 21 is a measurement of a fast relay. One curve designated x , is the displacement versus time. The other curve marked $\dot{x}$, is the velocity. In neither the displacement curve nor the velocity curve before impact, do the rates approach the overload condition.

The impacts are shown on the velocity curve to have the extreme rate of change. Such abrupt curves cannot be taken as being literally true. Following such abrupt changes the 10-kc oscillations also are found, but have not been drawn as they have to be discarded.

Discontinuity Errors Due to Averaging

Another effect acts to limit the sharpness of measurement when discontinuities occur. This is due to variations in successive operations of the device being tested. Obviously, if the actuation time varies over a range of ± 0.1 milliseconds, then any device such as the present one which measures by averaging many measurements, will not measure accurately in the close vicinity of such a discontinuity.

For a displacement measurement made at the average time of armature impact, part of the measurements will be before impact, the others after. The meter therefore will register too small a displacement. For another measurement a short time earlier or later than the impact part will be before and part after the average instantaneous displacement but the instrument average will still be good because the function now is smooth. The only displacement error then is just at impact and when the data are plotted, appears as a rounding of the curve. A good correction can be made merely by continuing the adjacent slopes to a sharp intersection.

A more striking effect is observed for velocity. At the moment of armature impact against core, the velocity drops suddenly from maximum to zero. The indication shown by the instrument then for this averaging type error is just half the final velocity. Another measurement made a short time earlier is good and the slope from that region can be extrapolated to the instant of the half velocity indication as the plot for the true velocity.

Choice of Output Transformer Inductance

The output transformer primary inductance is a compromise between output amplifier gain limitation due to equivalent shunt capacity of the inductor and the lowest desired cycle time. It was shown above that at 0.1 second cycle time only 1 volt is needed for a full scale deflection whereas 5 volts is available. As the instrument current is inversely proportional to the cycle time, 1.0-second cycle time will provide about half scale. The 11.5H chosen then, just barely covers the timing range of present interest. The impedance of the equivalent shunt capacitor at 10 kc is approximately equal in magnitude to the transformer dc resistance, and full internal gain is available over the operating frequency band.

The delay time is not altered by a change in inductance. For instance, if the inductance were halved the rate of current rise would be doubled but twice the change would be necessary for the same instrument deflection. The same result can be had merely by reducing the gain by one-half.

If longer cycle times were needed without the short times of present interest, the inductance could be appropriately increased, and the amplifier bandwidths correspondingly decreased, without loss of accuracy.

Summary of Amplifier Discussion

The foregoing discussion of the amplifiers covers the design considerations, the required frequency responses, and errors because of divergence from the ideal. For displacement measurements, all events occur in intervals greater than the minimum which the amplifiers can follow. This is also true of velocity measurements as regards the photocell amplifier and differentiator. The output amplifier does overload momentarily at jump discontinuities of velocity, and in the time vicinity following such events is in error.

Another type of error caused by variations in successive operations of the device being tested can be corrected by use of the measured rates just before and after the discontinuity.

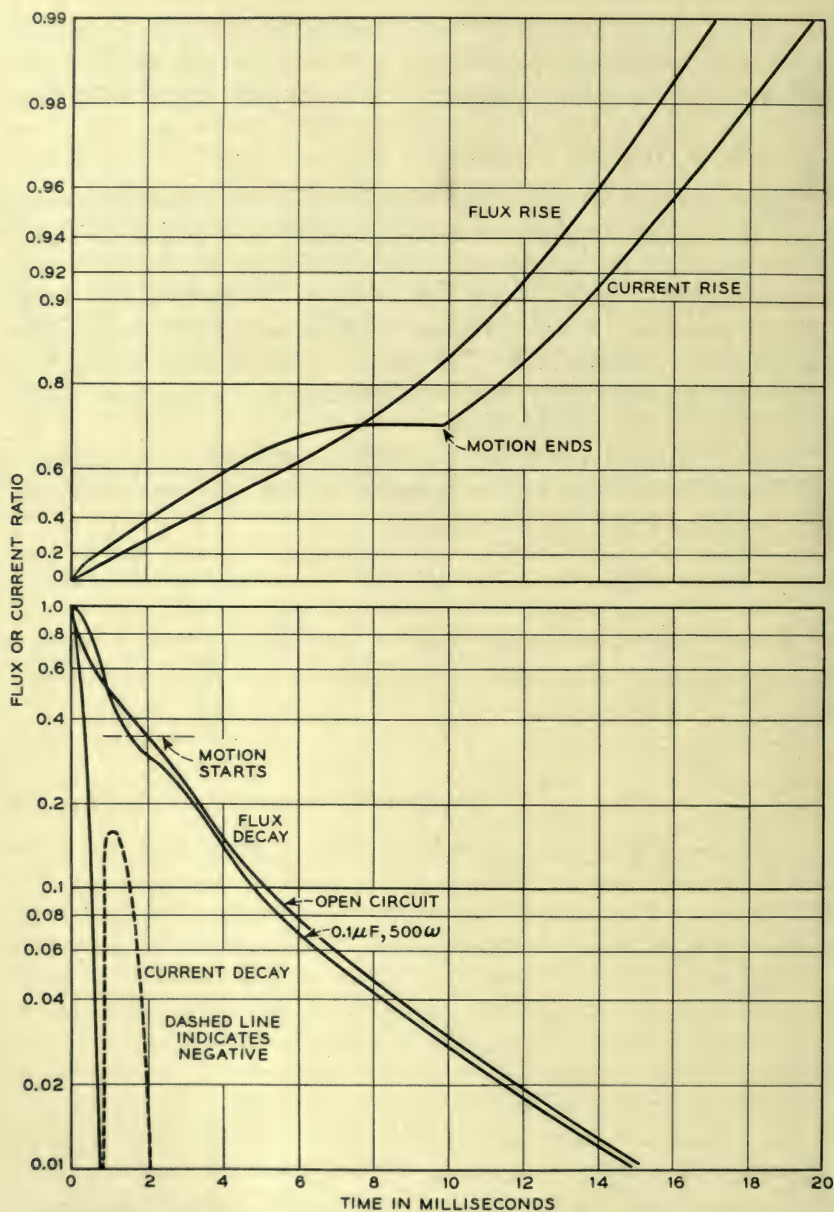


Fig. 22 — Winding current and flux measurements of an AF-type relay.

Part III — Applications

NEED FOR DYNAMIC FLUX MEASUREMENTS

The results of an experimental determination of dynamic flux rise and decay in solid core electromagnets are of general interest. Analytical solutions have been obtained for linear infinitely long rods or toroidal shaped structures where geometric simplicity exists. These solutions are in infinite series form. These solutions show that an elementary perfect representation can not be expected, even for these simple cases. Furthermore, an electromagnet has other factors which make any attempt at an analytic solution impractical. Some of these are:

- (a) The magnetic material is non-linear.
- (b) The flux density is non-uniform because of leakage flux.
- (c) The varying geometry of the magnetic parts, including the necessary gaps, make the boundary value problems unmanageable.
- (d) Motion of the armature during operate and release.

These all lead to the conclusion that a quick and accurate method of measuring dynamic flux changes is necessary for fundamental studies of the dynamic behavior of electromagnets.

As an example, Fig. 22 shows dynamic flux and current rise and decay measurements made on the same relay used for the dynamic motion studies. These data are plotted on semi-log graph paper as on such a plot an exponential curve becomes a straight line. The current rise curve shows the dip due to armature motion, ending abruptly when the motion is completed. The flux rise curve starts off nearly as an exponential but rises more rapidly after armature motion starts.

Two decay curves are shown, one with an open circuit and one with a contact protection network. Associated with the latter is the winding current which flows through the network. In the open circuit case, even without the effect of armature motion, the flux decay is not exponential. The flux decay with the network has a somewhat oscillatory shape about the open circuit curve. The current itself does complete one heavily damped cycle. The reversed current flow is shown as a dashed curve.

The decrease in flux after a short interval compared to the open circuit case, demonstrates that such a network can decrease the release time as well as afford contact protection.

For more fundamental studies, it is better to study flux behavior with the armature held fixed and avoid the motional effects.

DYNAMIC FLUX DECAY AND DEFINITION OF EQUIVALENT CORE CONDUCTANCE

Fig. 23 shows another measured flux decay curve with the armature locked in the operated position. Also shown are two one-term exponential

equations for comparison. For such an open circuit decay curve, the flux does not abruptly drop to zero because of induced core eddy currents. The analysis of these data thus partly resolves into the determination of some method for conveniently representing the eddy current effects.

NEW APPROXIMATE FLUX DECAY EQUATION

For a first approximation, use can be made of linear circuit theory. Assuming the core behaves as a coupled single turn of conductance G_e and inductance L_1 , and defining the core time constant as:

$$t_e = L_1 G_e, \quad (18)$$

the flux decay equation is the well known solution:

$$\frac{\varphi}{\Phi} = e^{-t/t_e} \quad t \geq 0. \quad (19)$$

It is clear that the dynamic flux decay cannot be represented by the above exponential equation, which has been drawn on Fig. 23 as the upper straight line.

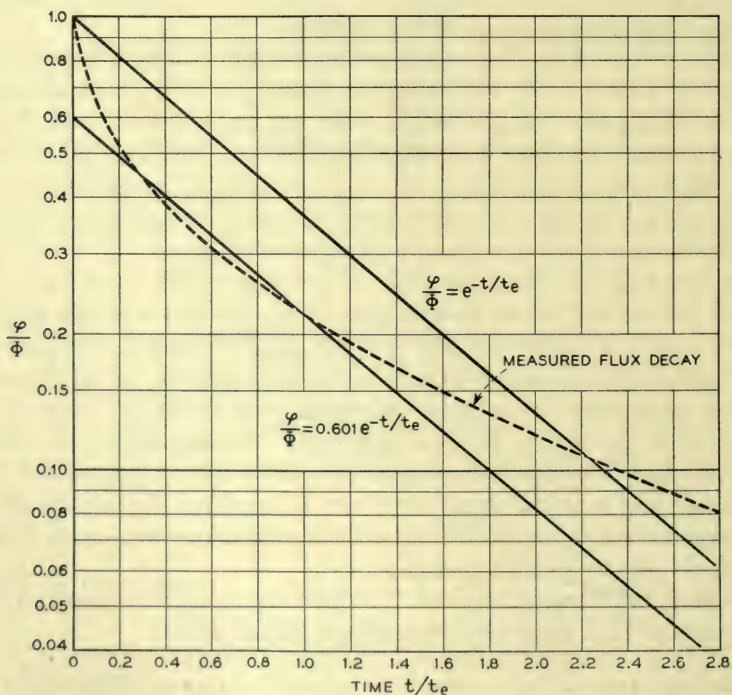


Fig. 23 — Comparison of measured open-circuit flux decay with one-term exponential equations.

The true decay curve has an initial steep slope followed by continuous curvature. A better approximation is to stipulate an initial jump discontinuity. A good fit to the flux range in which release of electromagnets occurs, shown by the lower straight line, is

$$\frac{\varphi}{\Phi} = .601e^{-t/t_e} \quad t \geq 0. \quad (20)$$

This discontinuity of flux in the first approximation is just the reverse of the continuity of flux concept usually used. Of course, no actual discontinuity occurs, as shown by the true decay curve. The failure of any single exponential equation to represent the true curve merely makes clear the fact that the behavior of the core is not that of a single coupled turn, but rather is that of an infinite line. Bozorth⁷ gives the intercept as 0.691 for the linear case, which does not represent the continuously curving decay which actually occurs.

The chosen intercept of the first approximation curve at $t = 0$ is admittedly somewhat arbitrary. It was arrived at in a broader study including flux rise curves. Accepting this equation, a convenient determination of t_e can be made similar to that for exponentials. If t is set equal to t_e , then

$$\left. \frac{\varphi}{\Phi} \right|_{t=t_e} = 0.221. \quad (21)$$

Thus, after measuring a dynamic flux decay curve, the time can be determined for which the above ratio obtains. This directly is t_e . From linear circuit theory and Lenz's Law, the inductance for one turn is:

$$L_1 = \frac{\Phi}{NI}, \quad (22)$$

whence G_e can be determined.

Now because of magnetic saturation and the shape of the hysteresis loop, the values will depend upon the particular final ampere turns (NI) used in the experiment. For uniqueness and uniformity in rating electromagnets for comparison purposes, the particular set chosen is that for which L_1 is a maximum. For comprehensive operating studies, measurements of course have to be made under the actual conditions of interest. Except for rating purposes, t_e is not ordinarily split up into components.

Thus, while the rated value of G_e for a particular electromagnet has the dimensions of conductance, it includes other factors as well. Some are: (a) Core material conductivity, (b) Magnetic non-linearity, (c) Shape of the hysteresis loop, (d) Non-uniform flux distribution, (e) Effect of pole

face and other air gaps, and (f) Decay of leakage field. It, therefore, also has the nature of a mop-up factor in which is included other effects not explicitly covered in the elementary analysis.

The above discussion has been directed toward establishing an easily determined core time constant having characteristics useful for rating purposes, and demonstrating that the distributed nature of the core eddy currents precludes an accurate representation of the core as a single coupled turn.

Open circuit flux decay measurements of round and rectangular solid core electromagnets of magnetic iron, 1 per cent silicon iron, and 45 per cent permalloy all are accurately represented by the measured flux decay curve of Fig. 23, after establishing appropriate values for G_e and L_1 in each case.

DYNAMIC FLUX RISE

Dynamic flux rise curves, measured with the relay armatures blocked open, are shown in Fig. 24. A family of curves, is needed because of the added variable of the winding. The value of t_e is determined by the method just described, except that the armature is held open, resulting primarily in a new and lower value of L_1 . The winding may be characterized by its coil constant:

$$G_c = N^2/R,$$

where N = number of turns, and R = dc resistance of winding circuit. These curves are a composite of measurements on many structures and fit all data to within a few per cent.

An empirical closed form expression has been determined which fits these data, and in fact was used to compute these curves. However, once the curves have been plotted there is no further need for the rather complicated expression. Any particular rise curve desired can be interpolated from the drawing.

These curves again show, with windings of relatively low coil constants, a divergence of the dynamic flux rise from being a straight line. The particular curve $G_c/G_e = 0$ actually is a decay curve because such a rise measurement would involve the use of both an infinite voltage battery and an infinite resistance winding. Now for this case, as well as for all the others, the rise and decay curves differ because of the shape of the hysteresis loop, the decay curve persisting longer. This effect is most evident for the last few per cent of the flux change. However, the differences are small on a full range basis. The curves shown near $G_c = 0$ are

a compromise for this effect. Actually the rise curves do not cross, but they do approach each other.

Coil to core constant ratios now in use range from 2 upwards, the case of 2 being the best condition for 25 watts power. An examination of this particular curve shows some curvature for small times. This effect of an initial increased rate of rise reduces the operate time of an electromagnet, and is automatically taken advantage of in the experimental design of windings.

With a winding as part of the system, an elementary consideration shows why this curvature exists. This follows from

$$N \frac{d\varphi}{dt} = E - iR, \quad (23)$$

where: N = number of winding turns, E = applied battery voltage, R = resistance of winding, i = instantaneous current, and φ = instantaneous flux. This equation is exact and holds whether or not there are eddy currents. At the instant of circuit closure there are no winding or eddy currents (because of leakage inductances if for no other reason) and the initial rate of flux rise is dependent only upon the number of turns and the applied voltage. After a transition interval, the eddy currents become effective and slow down the rate of flux rise. For any given total flux, the time required, therefore, is less than that given by the well known equation:

$$\frac{\varphi}{\Phi} = 1 - e^{-t/L_1(\sigma_c + \sigma_e)}. \quad (24)$$

However, for large coil constants where G_c predominates the effect diminishes and the above equation is an excellent representation and desirable for its simplicity.

NEW FIRST APPROXIMATION FLUX RISE EQUATION

Returning to Fig. 24, while the effect we have been discussing is all important for open circuit flux decay, practical coil constants at present do not have ratios to the core constant much below 2. These curves do not diverge greatly from straight lines on this plot. Can a minor correction term be made a part of the simple equation heretofore used which will retain its simplicity and extend its accuracy to windings now used?

One such equation is

$$\frac{\varphi}{\Phi} = 1 - e^{-t/L_1(\sigma_c + \sigma_e e^{-\sigma_e/\sigma_c})}, \quad (25)$$

where Φ is the final flux. This may also be written in integral form as

$$t = (G_c + G_e e^{-G_e/G_c}) \int_0^\varphi \frac{d\varphi}{NI - Ni} \quad (26)$$

The change is the introduction of the exponential modifying the core constant G_e .

Fig. 25 shows comparisons of curves for four different coil to core constant ratios. In each case, the actual flux rise, the older equation, and the new first approximation are shown.

For large coils (as an example on Fig. 25, $G_c/G_e = 10$) the actual flux rise is well represented by either expression, their accuracy being within 1 per cent.

For very small coils (on Fig. 25, $G_c/G_e = 0.5$) it is clear that the older representation is never a good approximation of actual flux rise. The new approximation represents the start of the dynamic flux rise quite well, but is not accurate for the second half. However no single exponential

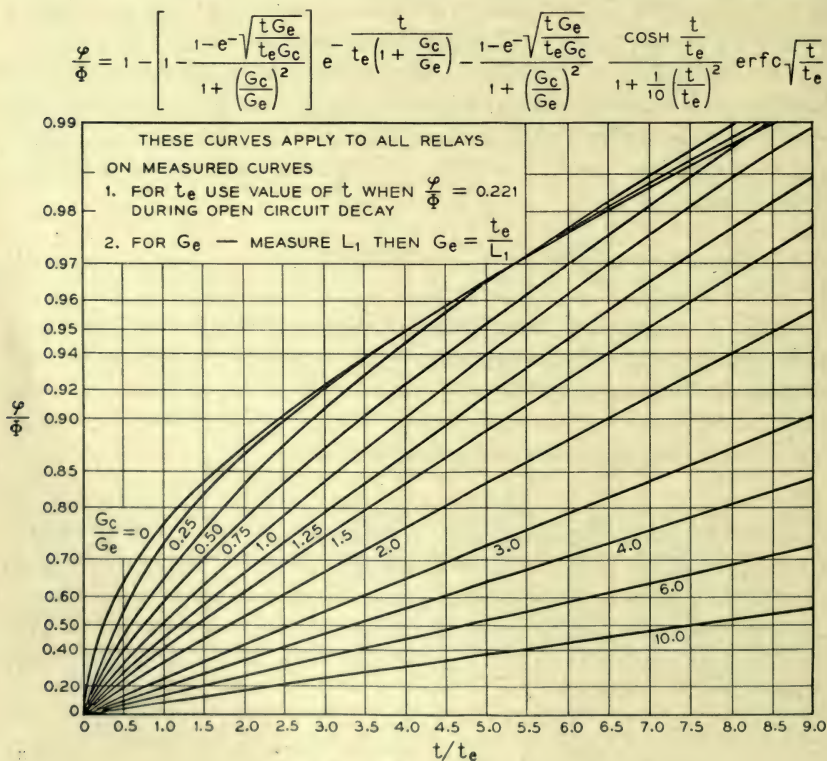


Fig. 24 — Flux rise curves.

equation can represent the actual rise when the plot is as curved as this one is. The best that can be done is to approximate the more important initial part with a straight line, and in this sense the new approximation shows a good fit.

A typical speed relay will have a value of G_c/G_e around 2 and the range from 1 to 10 covers all relays in this class. For a ratio of 2, the new representation of flux rise is within 2 per cent of the actual rise, as opposed to about 5 per cent with the earlier equation. For the ratio of 1, the accuracies are 4 per cent and 11 per cent respectively.

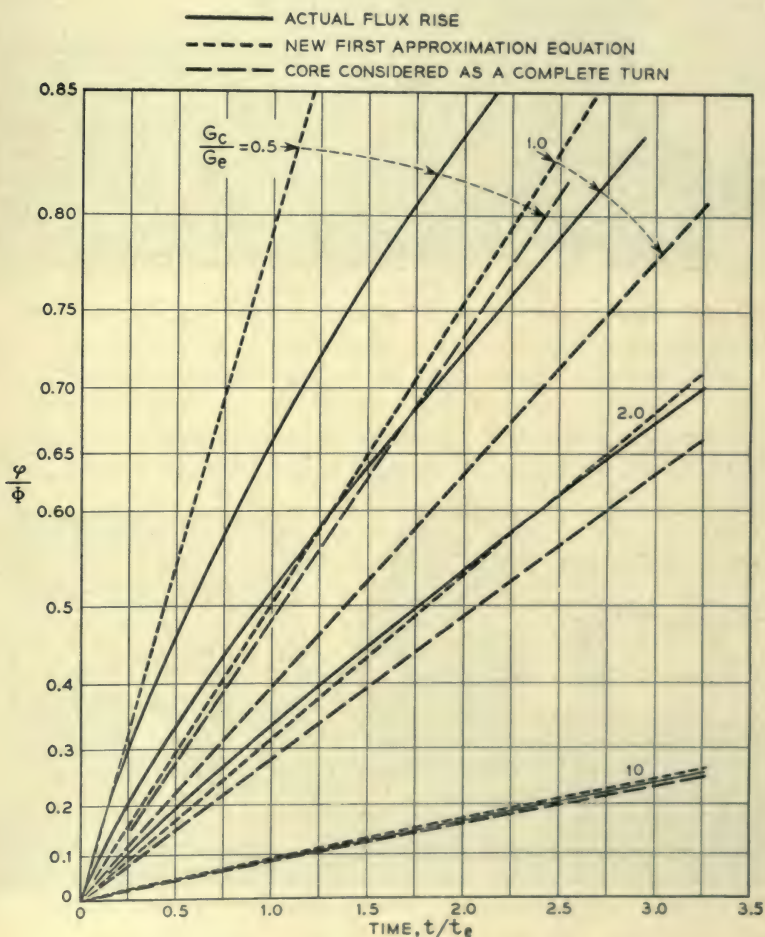


Fig. 25 — Comparisons of equations for flux rise.

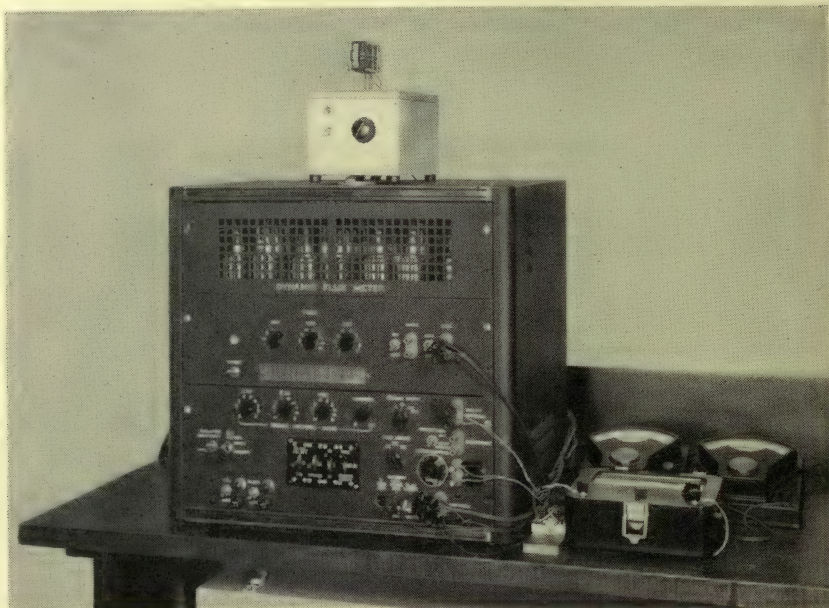


Fig. 26 — The dynamic fluxmeter.

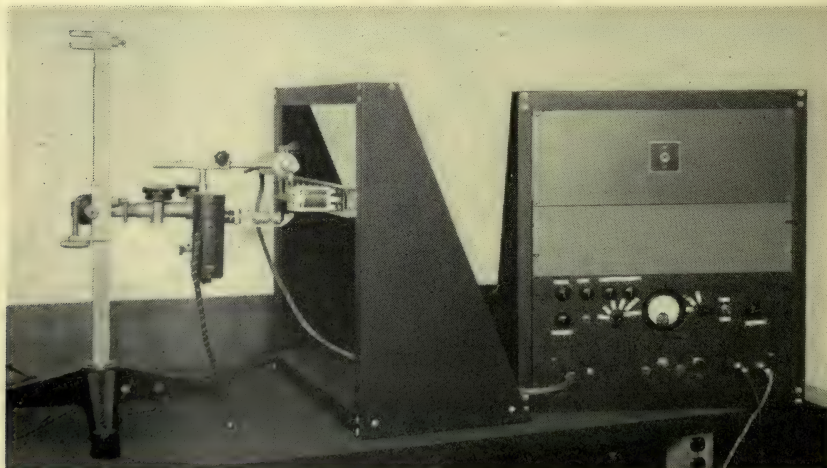


Fig. 27 — The optical probe and associated dc amplifier system.

Thus we have shown the new expression more accurately represents the eddy current effect on the initial flux rise for speed coils, and for most relays the overall accuracy will be within 2 per cent, and all relays should be within 4 per cent, an error reduction of about two-thirds compared to the older approximation.

EQUIPMENT

The dynamic fluxmeter is shown in Fig. 26. Two panels, each with a subchassis attached, comprise the set. The dc instrument and a supplementary voltmeter and ammeter are on the bench beside the cabinet. The upper panel contains the timing system, with the time selection controls. The lower panel includes the power supply and the test relay circuit controls, including the mercury contact relays. The control keys arrange the test relay circuit for any test condition.

The optical probe, a test relay, and the dc amplifier system are shown from left to right in Fig. 27. The photocell is in the shielded container above the relay, with the lamp and lens system below. A right angle prism turns the light beam into the vertical, to pass between the vanes on the relay.

The upper panel of the bench rack is the -250 -volt supply. The lower panel includes the $+250$ -volt supply, the three dc amplifiers and the magnetically shielded air-core output transformer. Its secondary is connected to the fluxmeter through a shielded cable. The controls are for zero setting the dc amplifiers, for selecting the amplifier gain and whether a displacement or velocity measurement is to be made.

REFERENCES

1. E. L. Norton, Dynamic Measurements on Electromagnetic Devices, A.I.E.E. Trans., **64**, p. 151, April, 1945.
2. Keister, Ritchie and Washburn, *The Design of Switching Circuits*, D. Van Nostrand, 1951.
3. Otto Schmidt, Thermionic Trigger, J. Scientific Instr., **15**, p. 24, 1938.
4. Richard Weissman, Stable Ten-Light Decade Scaler, Electronics, **22**, May, 1949.
5. M. Artzt, Survey of DC Amplifiers, Electronics, Aug. 1945.
6. Bode, *Network Analysis and Feedback Amplifier Design*, D. Van Nostrand 1945, p. 162.
7. R. M. Bozorth, *Ferromagnetism*, D. Van Nostrand, 1951, Chapter 17.

Selenium Rectifiers—Factors in Their Application

By J. GRAMELS

(Manuscript received July 1, 1953)

Selection of the proper selenium rectifier stacks for best results in the design of dc power supplies involves consideration of characteristics not ordinarily found in published data. This paper describes the data required for the selection of selenium cell sizes and cell combinations, shows typical voltage-current characteristics, and gives the results of extensive life test data necessary for evaluating the life expectancy of the product. The life test data indicate that there are substantial differences in the life expectancy of selenium stacks as manufactured by various companies in this country. Shorter life can be anticipated as the rms cell voltage ratings are increased. In addition, the life is affected by the current density and the temperature at which the selenium cells are operated.

INTRODUCTION

Since their introduction in this country about 1939, selenium rectifier stacks have proved to be a useful means of converting ac power to dc power for Bell System applications. These applications vary from 2 to 2,500 volts and in power sizes from a few watts to 10 kilowatts. Properly designed, the selenium rectifier has a relatively long-life expectancy and requires a minimum amount of maintenance. For these reasons, selenium rectifier stacks are widely used in telephone plants for battery charging, relay operation, plate and filament supply for vacuum-tube amplifiers, bias supplies, telegraph and teletypewriter circuits.

Up to 1952, about 245,000 rectifiers of all types have been manufactured for the Bell System. These include tungar, copper-oxide, vacuum tubes, thyratrons and selenium types. Of this total, about 25 per cent are of the selenium type.

Although Bell Laboratories studies of selenium rectifier stacks date from 1939, rectifiers using such stacks did not enter the telephone plant until 1945. From 1940 to 1945, however, selenium stacks were designed

widely into communication systems for World War II military projects. Since 1945, selenium rectifier power supplies have increased rapidly. For example, in 1945, out of a total of 9,000 rectifiers of all types, about 1,000 or 11 per cent used selenium. In 1951, about 30,000 out of 45,000 rectifiers, or 67 per cent, were of the selenium type.

There are many companies in this country who manufacture selenium stacks but the quality and behavior of the various manufacturers' products show considerable variation, particularly in regard to life expectancy. For many years the Laboratories has carried on an extensive testing program to evaluate properly the various suppliers' rectifiers. As a result of these continuing investigations, the Laboratories is in a position to select cell sizes and combinations of selenium rectifier stacks for use in new power applications. Specifications then are written on one or more suppliers. These specifications have a three fold purpose: (1) They are used as a purchasing and inspection document, (2) they cover the electrical and mechanical requirements necessary for proper electrical design and equipment layout, and (3) they are useful in maintaining records of the electrical and mechanical characteristics.

CELL MANUFACTURE AND STACK ASSEMBLY

A selenium rectifier cell is an elementary rectifying device having one positive electrode, one negative electrode and one rectifying junction. Cells are made by coating a chemically treated base plate, usually aluminum, with a thin layer of purified selenium to which a halogen element (chlorine, iodine or bromine) has been added. This mixture is applied to the base plate by one of the following methods:

1. Sprinkled or dusted on and subjected to heat and pressure.
2. Deposited by an evaporation process in an evacuated chamber.
3. Dipped in molten selenium and spun.

The selenium then is converted to the desired crystalline structure by heat treatments. During the heat treatment, a blocking or barrier layer is formed on the exposed surface of the selenium. This layer is further developed by various chemical means, which are "trade secrets" with each manufacturer. A thin layer of low melting point alloy, the front electrode, is then sprayed on the selenium.

The manufacturing process is completed by electrically "forming" the cells by applying a pulsating dc voltage in the non-rectifying direction for a specified time interval.

Selenium cells are assembled on an insulated bolt or stud, with individual cells separated by metal spacer washers to allow free passage of air for cooling the assembly. Contact terminals are brought out in various

arrangements for series, parallel or series-parallel connection of the cells as required.

The completed assembly is designated as a rectifier stack. A rectifier stack is a single structure of one or more rectifier cells.

Small size cells (less than 1") have no center hole for mounting. These cells are assembled without spacer washers in glass, metal, phenol fibre or bakelite tubing. Terminal leads extend outside these enclosures.

SYMBOLIC NOTATION

The combination of cells on a stack is described by a sequence of four symbols written a-b-c-d with the following significances:* (a) number of rectifying elements; (b) number of cells in series in each rectifying element; (c) number of cells in parallel in each rectifying element; and (d) symbol designating circuit or stack connections.

The symbols for the more common types of stack assemblies are shown schematically in Fig. 1.

A *basic* selenium stack is defined as a stack having a single selenium cell in each rectifying element. For instance, a 4-1-1B stack is a basic full-wave single-phase bridge rectifier stack with one selenium cell in each of the four rectifying elements.

The total number of selenium cells on any stack is the product of the three numbers indicated. For example, a single-phase full-wave bridge stack with three cells in series and two cells in parallel per rectifying element would be designated as a 4-3-2B stack assembled with $4 \times 3 \times 2$ or 24 cells.

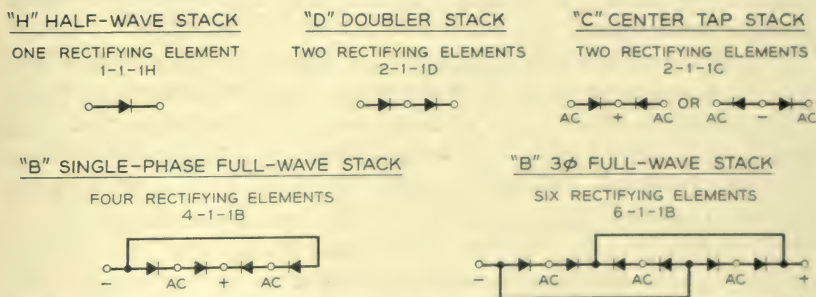


Fig. 1 — Stack assembly symbols. Color coding of terminals (AIEE and NEMA Standard): yellow for ac, red for plus dc output, and black for negative dc output.

* This method of specifying stack assemblies has been standardized by both the National Electric Manufacturer's Association and the American Institute of Electrical Engineers.

DESIGN CONSIDERATIONS

The proper selection of selenium stacks for use in dc power supplies involves a number of important factors that must be carefully considered.

(1) Circuit requirements must be carefully analyzed so that allowance may be made for the normal variations in the voltage-current characteristics of each manufacturer's product as well as variations that exist for the same stacks processed by different manufacturers. For a fixed ac input voltage, differences in the forward voltage drop may vary the dc output voltage at least ± 3 per cent from the mean value. If this cannot be tolerated, selenium cells have to be carefully graded and selected to obtain uniformity or other circuit adjustments have to be provided. Special selection of cells, obviously, will increase the cost of the product.

(2) The engineer must take into account the magnitude of the changes in the voltage-current characteristics of the rectifier stacks over the specified temperature range of his project. At very low temperatures, output voltages may be 5 to 10 per cent lower than at normal room temperatures. For high temperature operation, the stacks must be properly derated for both current and voltage to prevent overheating and rapid failure.

(3) Selenium rectifiers age with time. Compensation for this aging should be provided if load requirements warrant. The project engineer should determine what life he expects or requires of the application. For military applications life requirements may vary from minutes to thousands of hours. On other applications, such as telephone and elevator installations, it is desirable to design selenium rectifiers for life expectancies of ten to twenty years or more.

(4) The equipment engineer must anticipate the differences in mechanical details of the same stack assembled by different suppliers. There is no standardization in the selenium industry regarding cell sizes or mechanical details such as the overall length and height of the stack and particularly the type of mounting. However, a committee for the National Electrical Manufacturers' Association is attempting to standardize these mechanical details so that stacks assembled by different suppliers will be mechanically interchangeable.

(5) Unless otherwise specified, rectifying stacks are coated with various types of paints and varnishes for protection against moisture in normal conditions of humidity. For military projects and other applications where selenium rectifiers may be exposed to high humidities, fungus, salt or other corrosive atmospheres, the rectifier stacks must be

provided with a more suitable type of protective coating or finish. These finishes are available from most manufacturers.

(6) When selenium stacks are mounted in cabinets or housing with other heat-generating devices, they should be arranged in such a manner that heat from the other components does not reach the rectifier stacks. The stacks should be mounted below the other components, and in such position that the free flow of air through the rectifier stack is not impeded. The stack should be mounted with the assembly stud in the horizontal, not vertical, position. When two or more stacks are mounted in the same housing, they should be in a horizontal plane with each other, or properly staggered. Cabinets should be provided with louvers to dissipate the heat within the enclosure.

ELECTRICAL RATINGS

The electrical ratings of selenium rectifier stacks are based on their voltage, current and thermal characteristics. All three must be considered carefully for initial design purposes, as any one can affect life expectancy.

Voltage Ratings

The voltage rating usually is expressed as the "reverse voltage rating." It is the maximum rms sinewave voltage above which an excessive reverse current would flow and overheat the cell, causing breakdown. However, the important consideration establishing the rating is the peak voltage applied to the cell. If the applied voltage differs significantly from a sine-wave, it is important that the applied peak voltage shall not exceed 1.41 times the rated rms voltage.

When selenium cells originally were manufactured in this country, their rms reverse voltage ratings ranged from 14 to 18 volts. Ratings later increased to 26 volts. One manufacturer already has successfully produced 33-volt cells for several years; and another supplier announced recently a 40-volt cell. Cells rated at higher voltage have been produced in the laboratory. For non-critical applications, such as low-cost radio and television sets, some selenium manufacturers make cells rated at 45 volts rms. Generally, in such applications, a high reliability product is not required and the units may have a relatively short life.

The nominal dc output voltages obtained from various common *basic* rectifier stacks assembled with cells rated at 18-, 26- and 33-volt rms are listed in Table I. These ratings apply for stacks operating at rated dc output current into a resistance load.

Current Ratings

The current rating of selenium cells is based upon a normal current density of 0.32 ampere for each square inch of active rectifying surface for a single-phase full-wave bridge rectifier stack (4-1-1B) operating into a resistance load. The rating applies to stacks in which the cells are separated by spacer washers so that the cells are cooled by convection air currents. Small size cells mounted without spacer washers are rated at a lower current density.

The dc output current ratings for various selenium cell sizes are listed in Table II.

These ratings are based upon continuous operation in ambient temperatures up to $+35^{\circ}\text{C}$ with unrestricted ventilation for convection cooling of the stacks. For *battery or condenser loads* on single phase circuits, the above current ratings usually are derated to 80 per cent of the above values.

Thermal Ratings

Voltage and current ratings usually are based upon operation of the rectifier stack in a normal ambient temperature of $+35^{\circ}\text{C}$. It is important to note, however, that for selenium cell application, ambient temperature is defined as the temperature *immediately* surrounding the stack within the equipment enclosure, not the temperature area where equipment is installed. Stacks operating at rated current and voltage into a resistance load have temperature rises ranging from 15 to 30°C when measured with a thermocouple in direct contact with the cell. The temperature rise depends upon the manufacturing techniques used by the various companies as well as the spacing of the selenium cells on the assembled stack. For long life expectancy, the actual cell temperature should not exceed $+75^{\circ}\text{C}$.

TABLE I—NOMINAL DC OUTPUT VOLTAGES OF RECTIFIER STACKS

Cell Rating, Volts rms.....	Basic Stack	DC Output Volts		
		18	26	33
Single Phase				
Half-wave.....	1-1-1H	6	9	12
Full-wave				
Center tap.....	2-1-1C	6	9	12
Bridge.....	4-1-1B	12	18	24
Three-Phase				
Full-wave bridge.....	6-1-1B	19	30	38

For operation in ambient temperatures above $+35^{\circ}\text{C}$, the voltage or current ratings, or both must be reduced to limit the cell temperature (ambient plus heat rise) to $+75^{\circ}\text{C}$. The de-rating factors for voltage and current vary somewhat from supplier to supplier, but Fig. 2 shows typical de-rating curves for operation of selenium stacks above a $+35^{\circ}\text{C}$ ambient.

Selenium rectifiers can be operated at cell temperatures above 75°C , but aging accelerates with temperature and excessive temperatures will result in rapid failure.

A special problem today is the application of selenium rectifiers for high temperature military uses. Many military projects are requesting designs to operate at ambient temperatures of $+70$ to $+90^{\circ}\text{C}$. Unfortunately very little data about rectifier operation and life expectancy under these conditions have been obtained up to now.

VOLTAGE — CURRENT CHARACTERISTICS

To demonstrate the non-linear characteristics of selenium rectifier cells, Fig. 3 illustrates representative dynamic voltage-current curves showing the open circuit and short circuit characteristics of bridge connected stacks composed of various cell sizes rated at 26 volts, rms. This rating was chosen since all manufacturers presently are producing cells of this type.

Cells with higher voltage ratings generally have a slightly greater forward voltage drop.

TABLE II — DC OUTPUT RATINGS

Cell Size	Area Square Inches*	Continuous DC Amps at $+35^{\circ}\text{C}$		
		Single-Phase		Three-Phase
		Half wave	Bridge & C.T.	Bridge
$\frac{1}{4}$ " Dia.	0.049	0.006	0.012	0.018
$\frac{1}{2}$ " Dia.	0.196	0.025	0.050	0.075
1" x 1"	0.56	0.090	0.18	0.270
$1\frac{1}{4}$ " x $1\frac{1}{4}$ "	1.1	0.175	0.35	0.525
$1\frac{1}{2}$ " x $1\frac{1}{2}$ "	1.7	0.270	0.54	0.810
2" x 2"	3.1	0.500	1.0	1.50
3" x 3"	7.0	1.10	2.2	3.30
$4\frac{3}{8}$ " Dia.	12.5	2.00	4.0	6.00
5" x 5"	21.2	3.40	6.8	10.2
$4\frac{1}{4}$ " x 6"	21.2	3.40	6.8	10.2
5" x 6"	26.2	4.20	8.4	12.6

* Approximate active rectifying area (varies with suppliers)

The data shown in Fig. 3 and subsequent data are plotted for *basic* 4-1-1B stacks. Tests were made with 60-cycle sinusoidal supply voltage. The "forward voltage drop" is expressed as the rms volts required to produce a specified current in a moving coil dc ammeter connected directly to (short circuiting) the output terminals of the rectifier stack. For stacks with more than one cell per element, the voltage drop is obtained by multiplying the observed drop in Fig. 3 by the number of cells in series per rectifying element.

The reverse current is measured by applying a specified rms voltage to the ac terminals with the dc terminals open circuited and noting the rms input current after the current has stabilized. When selenium stacks have been "off voltage" for some time, a relatively high reverse current is obtained for the first few seconds. The current then decays approximately exponentially. Usually, the current will stabilize after voltage has been applied for 5 to 10 minutes.

It should be emphasized that these curves are only typical characteristics. Selenium cell manufacture requires individual testing of each cell before assembling into a stack. Cells are graded by their electrical characteristics. Large variations exist between the lowest and highest grade. Each manufacturer sets up his own standards regarding the variations

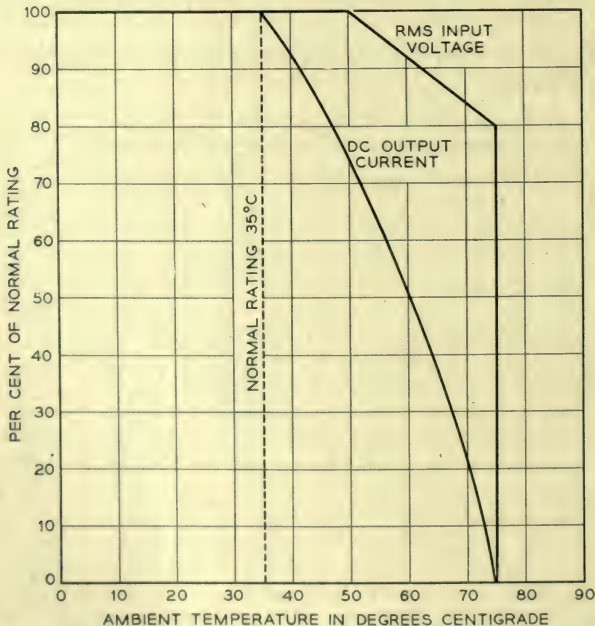


Fig. 2 — Temperature de-rating curves for long-life expectancy.

of characteristics in a given grade. More than one grade of cell may be assembled in a rectifier stack. The art of manufacturing selenium cells is such that, considering production over a yearly period, differences in the forward voltage drop at rated current may vary as much as ± 30 to 50 per cent from the mean value. In the reverse direction, a larger per cent spread exists in the reverse current, particularly in cells or stacks made by different suppliers.

Fig. 4 shows dynamic characteristics plotted on a linear scale to illustrate variations of the forward voltage and reverse current characteristics of selenium rectifier stacks that are processed by two different suppliers. Variation of this magnitude exist, not only from supplier to supplier, but also may occur in a particular suppliers' product.

Selenium rectifier stacks in common with other semi-conductor rectifiers, have a negative temperature coefficient of resistance in the forward direction. The forward voltage drop at a specified current decreases as the ambient temperature increases (see Fig. 5). In the reverse direction, the reverse current decreases as the temperature is lowered to approximately -20°C . Below this temperature, there is no apparent change in the current except at the higher voltages. At the higher voltages, the current again tends to increase.

STACK DESIGN

The dc output voltage-current characteristics for various ac input voltages of *basic* rectifier stacks (one cell per rectifying element) are represented in Figs. 6, 7 and 8. The data were obtained by maintaining a constant 60-cycle rms voltage at the stack input terminals and do not take into account transformer regulation or other regulating devices that may be used.

Fig. 6 shows the single phase full-wave bridge characteristics for *resistance* loads.

Fig. 7 shows the single-phase full-wave bridge characteristics for *battery* loads. It will be observed that, with single-phase circuits for a given dc output voltage, lower ac input voltage is required for a battery load than for a resistance load. Capacitor loading is somewhat similar in output characteristics to battery loading, but the value of output voltage is dependent upon the magnitude of the capacity, the quality of the capacitor, and the current drawn by the load. For battery loading, the output voltage is dependent upon the type of battery, the condition of the battery, the battery voltage and the charging current rate required.

For these reasons, it is difficult to accurately predict the exact input-

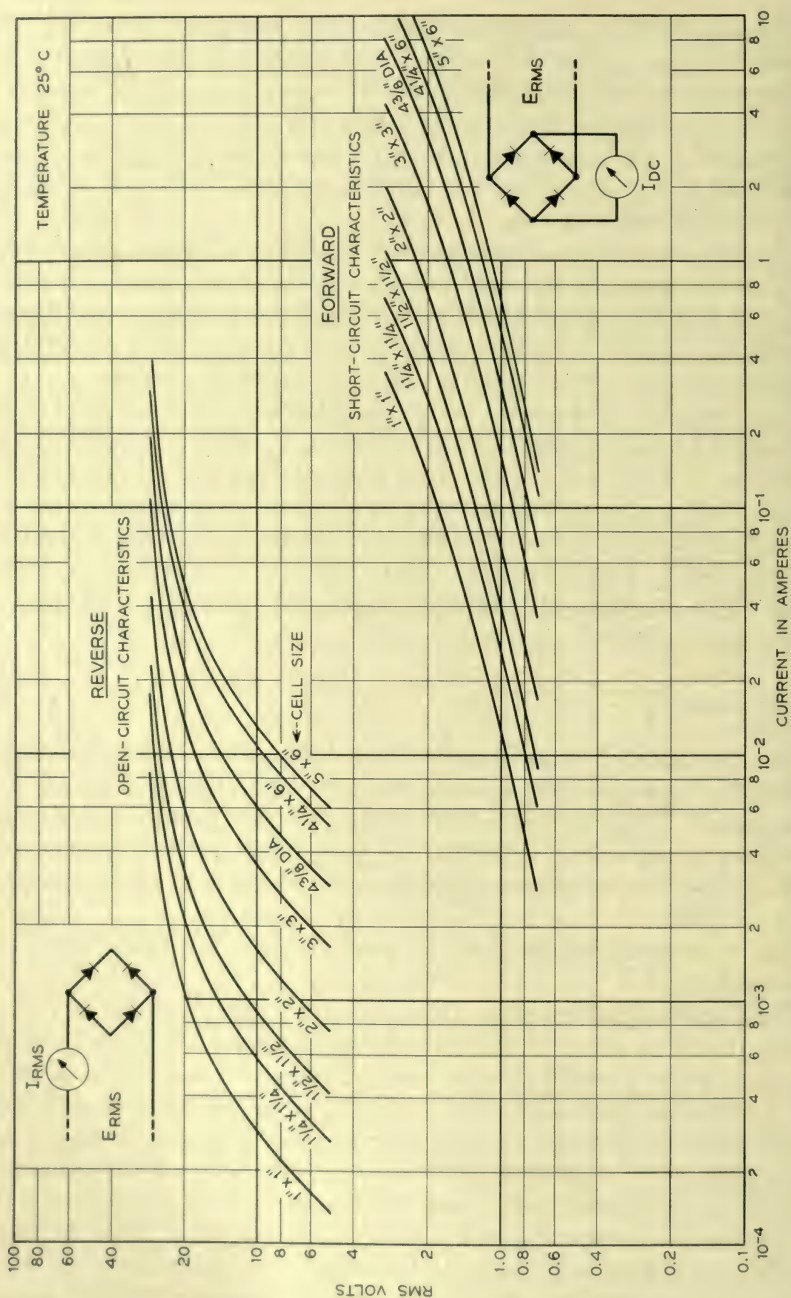


Fig. 3 — Typical dynamic voltage-current characteristics of 4-1-1B stacks with 26-volt cells.

output characteristics of stacks on capacity or battery loads. For precise design information, laboratory tests under the specified conditions should be made to determine the actual slope of these characteristics.

Fig. 8 shows the characteristics for a three-phase full-wave bridge circuit. These characteristics apply to both resistance and battery loads.

To design a selenium rectifier stack, the following procedure may be used:

1. Refer to Table I for the dc output voltage rating for the particular circuit application.
2. For the specified dc output current select the proper cell size from Table II. (If the current exceeds that of a single cell, additional cells may be connected in parallel.)
3. The ac input voltage to the stack for the required dc output voltage and current is then obtained from Figs. 6, 7 or 8.

The following example illustrates this method of stack design.

Example: To design a single-phase full-wave bridge rectifier stack to supply 1.0 ampere dc at 48 volts into a resistance load. From Table I, the highest dc output voltage for a basic stack is 24 volts for cells rated

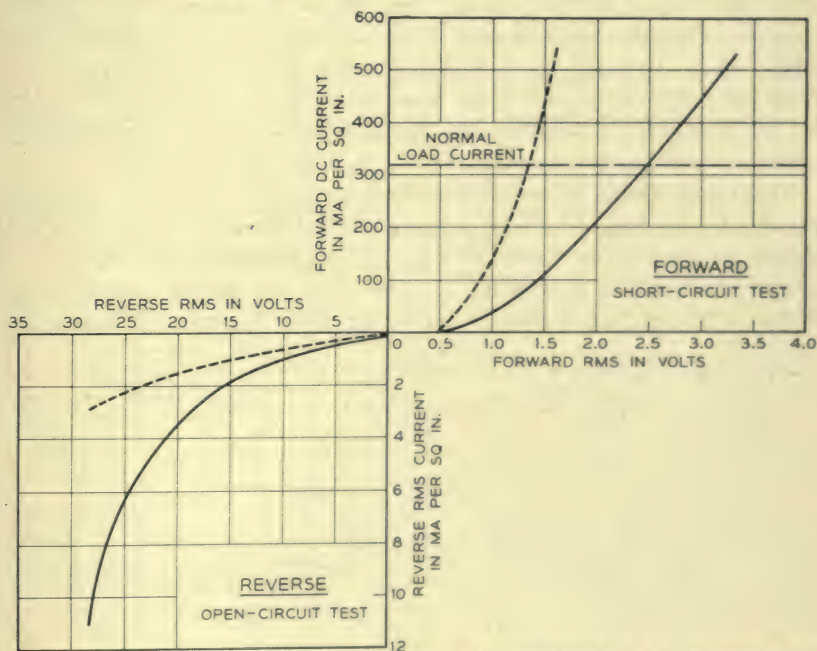


Fig. 4 — Typical dynamic voltage-current characteristics of 4-1-1B stacks with 26-volt cells processed by two different suppliers.

at 33 volts, rms. Therefore, $48\frac{1}{2}$ or 2 cells in series in each of the four rectifying elements of the full wave bridge circuit will be required. From Table II, a 2-inch square cell is needed to carry the specified load of 1 ampere dc. As previously described under "Symbolic Notation," the rectifier stack would be designated as 4-2-1B combination using 8 cells.

Selenium cells rated at 18 or 26 volts rms also could be used, but the total number of cells in the stack would be increased as shown in Table III.

Obviously, the higher the cell voltage rating, the less total cells and consequently a smaller and lighter stack is obtained. However, other factors, particularly life expectancy, must be evaluated as described later in the text.

After the stack cell combination has been established, the required ac input voltage is obtained from Fig. 6. At 100 per cent normal load current (1 ampere dc from Table II) and 24 volts dc, the ac voltage for a *basic* stack is about 30 volts, rms. Since there are two cells in series in each rectifying element, 2×30 or 60 volts rms input voltage would be required for a new stack.

If cells rated at 26 volts rms are selected, three cells in series would be required. For the specified 48 volts dc output, $48\frac{1}{3}$ or 16 volts dc per series cell, is obtained. Again referring to Fig. 6, an input voltage of 20.5 volts per cell or 61.5 volts total is required. This compares to 60 volts for the previous illustration. The higher input voltage results from the added voltage drop of the additional cells in the rectifier stack.

This stack design is satisfactory for operation in ambient temperatures not exceeding $+35^{\circ}\text{C}$. For operation at higher ambients, the stack should be de-rated as shown in Fig. 2. For example, if the stack is required to operate in an ambient of $+60^{\circ}\text{C}$, the dc output current rating would be 50 per cent of the normal rating. The 2" square cell previously

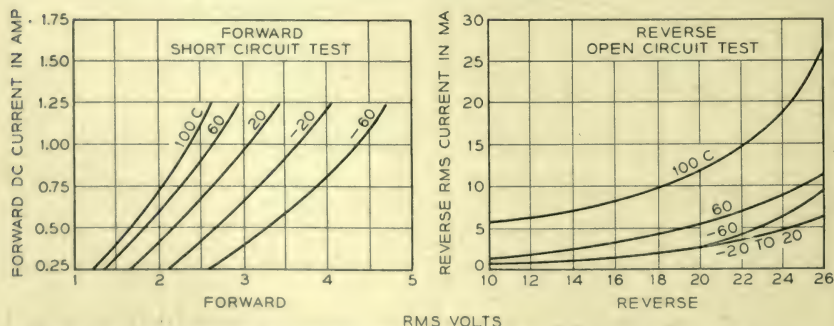


Fig. 5 — Temperature characteristics of 4-1-1B stacks with 2" x 2" cells rated at 26 volts rms and 1.0 ampere dc.

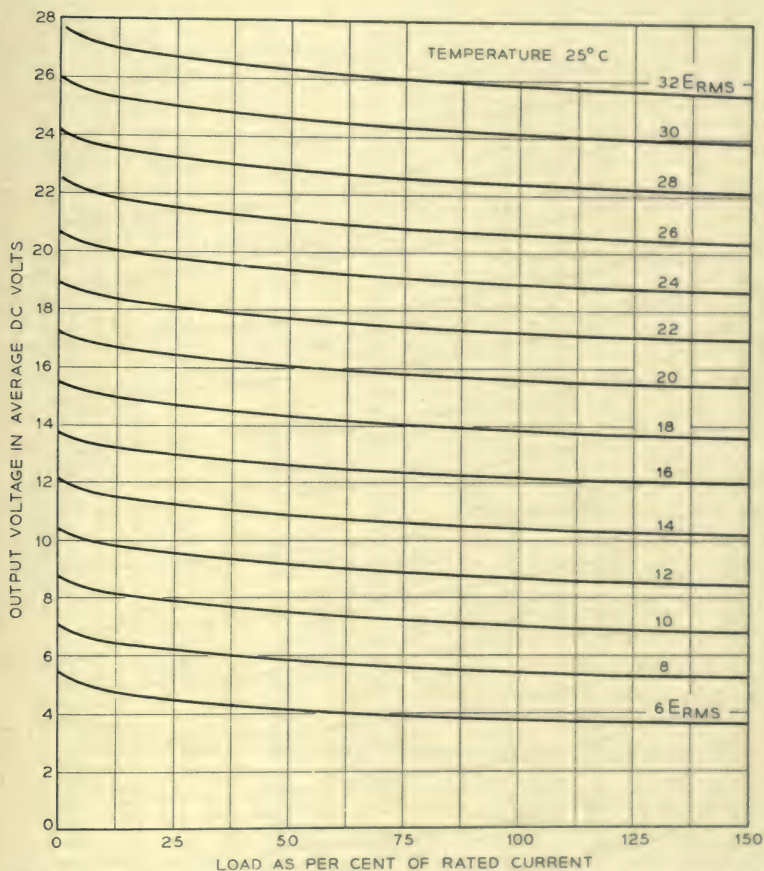
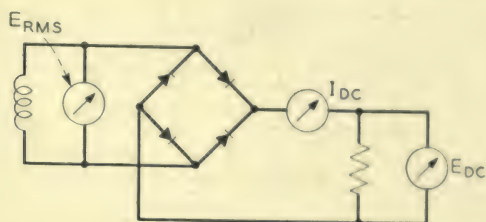


Fig. 6 — Typical output voltage-current characteristics. 60-cycle single-phase full-wave bridge circuit. Basic stack 4-1-1B. Resistance load.

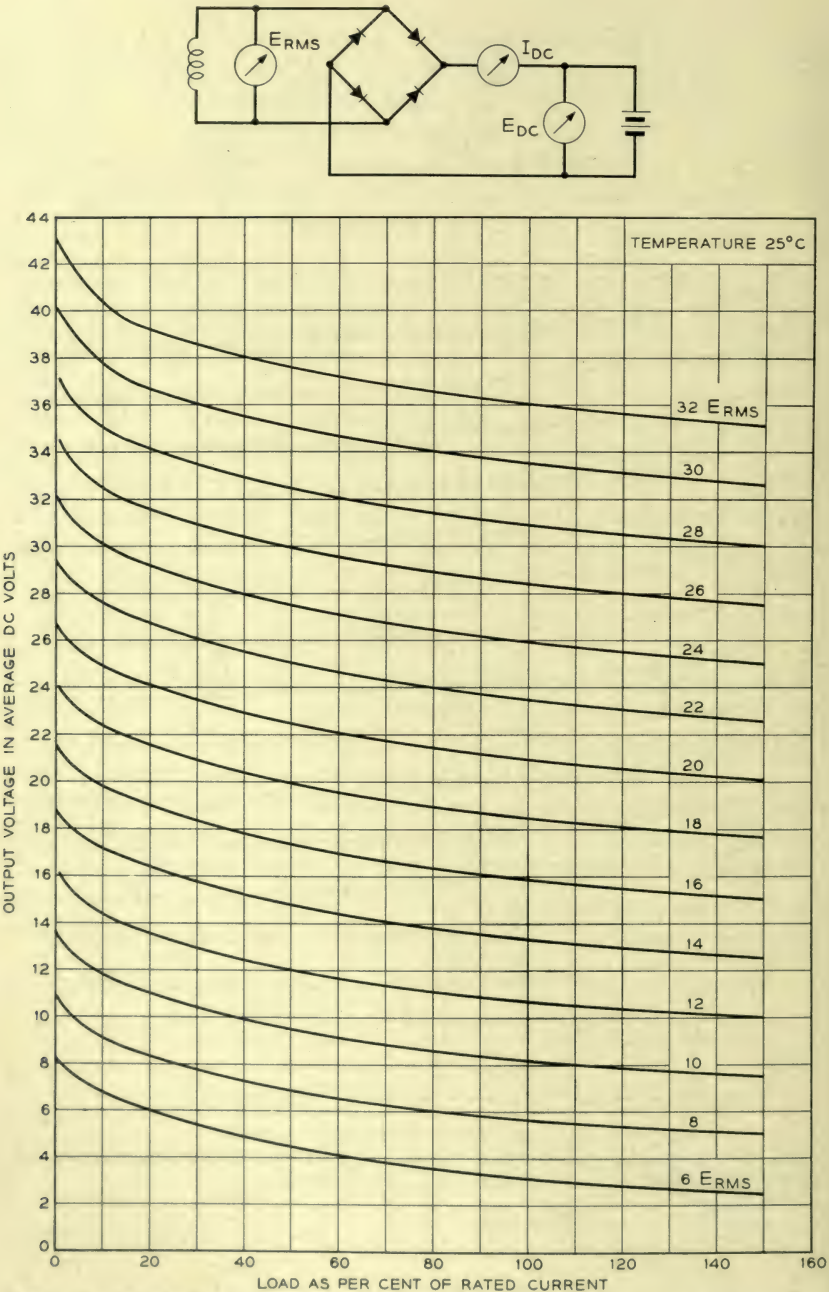


Fig. 7 — Typical output voltage-current characteristics. 60-cycle single-phase full-wave bridge circuit. Basic stack 4-1-1B. Battery load.

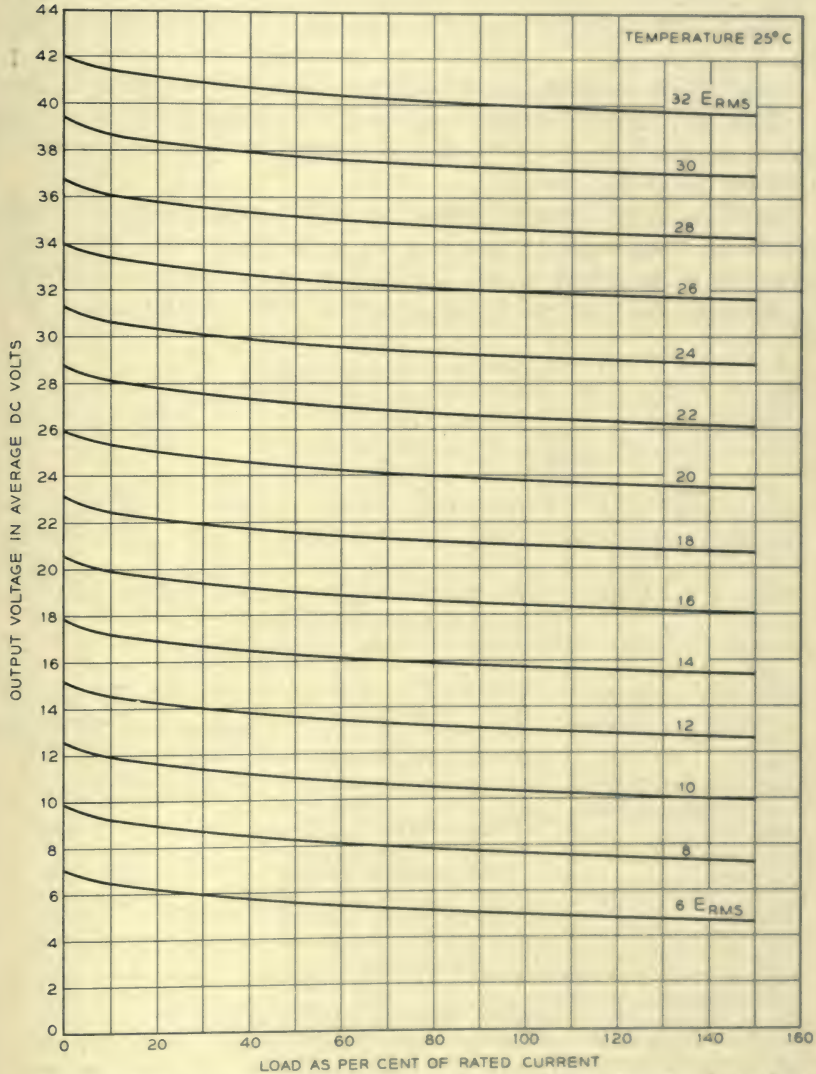
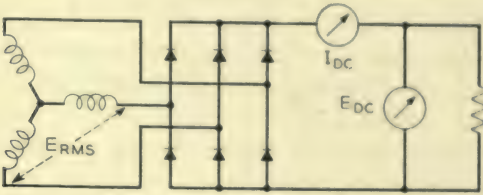


Fig. 8 — Typical output voltage-current characteristics. 60-cycle three-phase full-wave circuit. Basic stack 6-1-1B. Resistance and battery loads.

selected would now be rated at 0.50×1 or 0.5 ampere. Therefore, for the specified 1.0 ampere load, two cells in parallel are required or the next larger cell size, (3" square) could be selected. The 3" square cell would be rated at 0.50×2.2 or 1.1 ampere.

At 60°C ambient, the input voltage for new or aged stacks should not exceed 92 per cent of the normal rms voltage rating.

An alternative method of calculating the input voltage required for any stack cell combination may be obtained by the use of the formulas shown in Table IV.

Since the selenium rectifier stack is a non-linear device whose characteristics vary depending upon the circuit and operating conditions, the values given in Table IV are not precise but they are reasonably accurate for most design purposes.

To illustrate the formulas method, let us determine the input voltage for the previous example, that is, a single-phase full-wave bridge stack to supply 1.0 ampere dc at 48 volts into a resistance load. As previously determined, the stack cell combination would be a 4-2-1B circuit. The

TABLE III — STACK DESIGN

Cell Rating Volts rms.	Cell Combination	Total Cells
33	4-2-1B	8
26	4-3-1B	12
18	4-4-1B	16

TABLE IV — CALCULATING THE INPUT VOLTAGE REQUIRED FOR ANY STACK CELL COMBINATION

Rectifier Circuit	Formulae	K	
		Resistance Load	Battery or Capacity Load
Single-phase Half-Wave.....	$E_{ac} = KE_{dc} + nDV$	2.3	1.0
Single-phase Full-Wave, Center TAP*.....	$E_{ac} = KE_{dc} + nDV$	1.15	0.80
Single-phase Full-Wave Bridge.....	$E_{ac} = KE_{dc} + 2nDV$	1.15	0.80
Three-phase Full-Wave Bridge.....	$E_{ac} = KE_{dc} + 2nDV$	0.74	0.74

E_{ac} = Stack input volts, rms.

E_{dc} = Average dc output volts.

K = Circuit form factor.

n = Number of cells in series in each rectifying element.

DV = RMS forward voltage drop at specified dc output current (see Fig. 9).

* For center tap circuits, E_{ac} is the voltage to the mid-tap on the transformer.

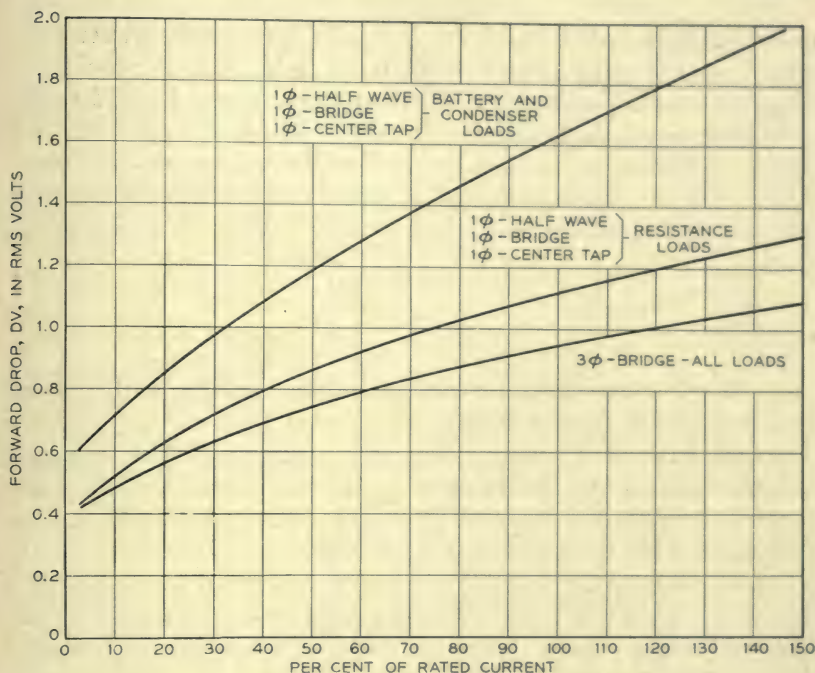


Fig. 9 — Dynamic forward voltage drop per cell.

forward rms voltage drop per cell (DV) obtained from Fig. 9 is 1.10 volts at 100 per cent normal rated current. From the formulas in Table III, the input voltage is:

$$E_{ac} = 1.15E_{dc} + 2nDV,$$

$$E_{ac} = 1.15 \times 48 + 2 \times 2 \times 1.1 = 59.6 \text{ volts rms.}$$

This input voltage is required for a new or unaged stack. Since the stack ages under service operating conditions, additional input voltage will be required to maintain the original dc output. It is considered good practice to design for a 100 per cent increase in the initial forward voltage drop (DV) of the stack. The aged stack, then would require

$$E_{ac} = 1.15E_{dc} + 2(2nDV),$$

$$E_{ac} = 1.15 \times 48 + 2(2 \times 2 \times 1.1) = 63 \text{ volts rms.}$$

The amount by which an increase of 100 per cent in the forward voltage drop or forward resistance will change the output is dependent upon the design of the circuit. It depends upon the ratio in per cent of the forward rectifier stack resistance to the total circuit resistance including

the transformer, ballast, load, etc. For most practical design considerations, one or more aging taps on the input transformer to provide 5 to 10 per cent additional voltage, will compensate for the forward aging.

It should be emphasized that the curves in Figs. 6, 7, 8 and 9 are based upon empirical data obtained on new rectifier stacks and the characteristics will vary slightly depending upon normal manufacturing variations in the voltage current characteristics mentioned earlier in this article.

AGING

Selenium rectifier stacks are subject to aging. Aging is defined as any persistent change (except failure) which takes place for any reason in either the forward or reverse resistance characteristics. The important factor in selenium rectifier aging is the increase in the forward voltage drop which results in a decreased dc output. For normal rectifier applications aging of the reverse current is not critical.

For design purposes, a selenium stack is considered to have reached the end of its useful life when the stack input voltage required to maintain rated output voltage exceeds the rms reverse voltage rating assigned by the manufacturer. Operation beyond this limit will result in overheating of the selenium cells and rapid failure of the stack.

The extent and rate of aging of selenium stacks cannot be predicted mathematically. Aging characteristics must be determined by actual test involving lengthy time consuming projects. To determine whether a given rectifier stack will give satisfactory performance for five years, as an example, tests would have to be conducted for this period. Aging can be accelerated by operation at high temperatures or at load currents above normal, but no accepted correlation exists between this type of aging and that obtainable under normal operating conditions.

Aging data were obtained on sample rectifier stacks obtained from different manufacturers in this country. The samples were set up on life test racks as single phase full wave rectifiers and were operated continuously at normal room ambient temperatures. For the duration of the tests, the 60-cycle ac input voltage to the stacks was kept constant at approximately 10 per cent below the maximum rms voltage rating. The stacks operated into a resistance load. The resistance was selected so that the rectifiers operated at rated dc load currents. The resistance was not changed thereafter.

Long term forward aging characteristics of selenium rectifier stacks

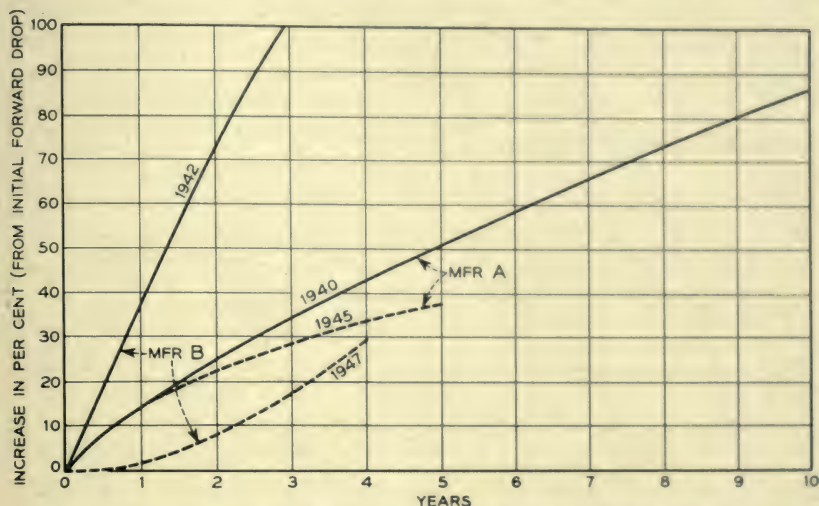


Fig. 10 — Forward aging characteristics of 18-volt cells.

assembled with cells rated at 18 volts rms are shown in Fig. 10.* After 10 years (approximately 90,000 hours) continuous operation, the forward drop on one supplier's product increased approximately 90 per cent. A second supplier's product increased 100 per cent in 3 years under similar operating conditions. Further development programs by each manufacturer resulted in improved aging qualities as indicated by tests made on stacks obtained at a later date.

By introducing changes in their manufacturing techniques, the industry eventually increased the voltage ratings to 26 volts rms per cell. This change in processing, however, raises the question — how has the operating life been affected? Fig. 11* clearly illustrates that although many manufacturers are commercially producing 26-volt cells, the life expectancy is vastly different.

In order to save space, weight and critical materials, manufacturers are attempting to produce cells with still higher voltage ratings. Most companies have had a 33-volt cell development program under way for some time, but again the problem of aging must be considered. Fig. 12* compares the forward aging of 26- and 33-volt cells after 5,000 hours operation. It is evident that in all cases except one, the 33-volt cells age substantially faster than the 26-volt cells. The exception is manufacturer "E" who has not produced 26-volt cells but has commercially processed

* The designations A, B, C, etc. shown on all aging curves are not to be interpreted to represent the same manufacturer's product on each of the figures.

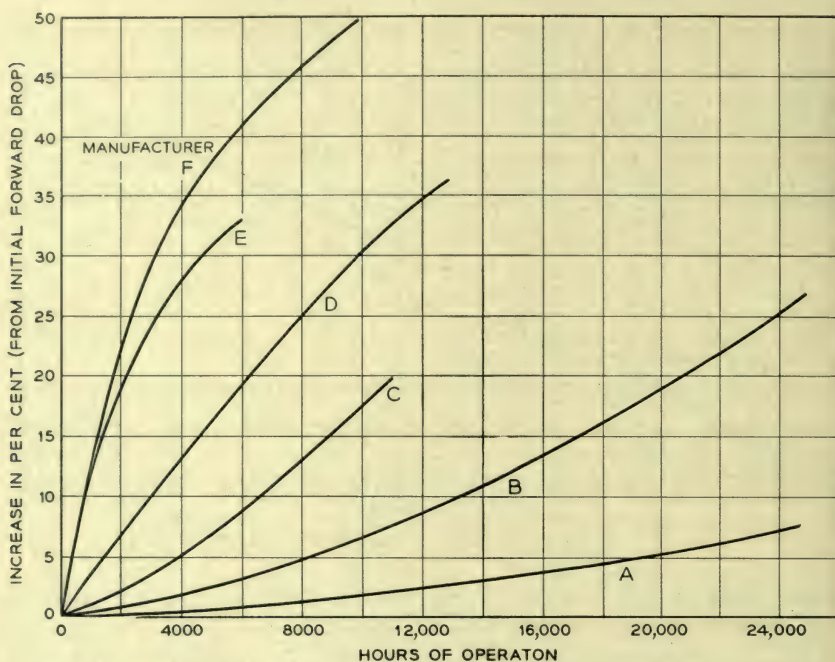


Fig. 11 — Forward aging characteristics of various suppliers 26-volt cells.

33-volt cells for several years. Again, there is a wide difference in the rate of aging in the various manufacturers' products.

Continued aging tests on stacks made by manufacturer "E" show that after 21,000 hours of continuous operation, the forward voltage drop has increased only 8 per cent. Extrapolation of these data indicate a life expectancy of thirty years.

It should be emphasized that the data on the 33-volt cells, with the one exception, were taken on experimental stacks obtained from various companies while they still were in the research and development stages of processing the 33-volt cell. Undoubtedly, further development programs on this type of cell will result in improved aging qualities.

Fig. 13 shows clearly what happens to the aging rate as manufacturers attempt to increase still further the voltage rating. These data are based upon 26-, 33- and 40-volt cells made by one supplier who is commercially producing cells with these ratings. The aging rate is accelerated greatly as the voltage rating is increased.

All the foregoing aging data were taken on stacks operated at rated load currents. A longer operating life may be expected if selenium stacks are selected to operate at lower load currents. As shown in Fig. 14, the aging

at 50 per cent normal current rating is about one half that at rated current. Above normal rating, the aging is accelerated to a greater degree. These data were obtained on a particular supplier's stacks with 26-volt cells, but similar behavior has been observed on other suppliers' products, including stacks assembled with 33- and 40-volt cells.

The behavior of the reverse current during these aging studies indicates that the reverse current generally decreases slightly during the first few months of operation and then remains substantially constant thereafter. A few manufacturers' rectifiers, however, show the opposite effect. The current rises slightly before leveling off. For normal dc power applications, these changes are insignificant.

As previously discussed, selenium stacks are rated for operation at $+35^{\circ}\text{C}$ ambient temperatures and should be derated at higher ambients for normal life expectancy. However, the problem frequently arises — what life can be expected at high ambient temperatures and why cannot aging qualities be predicted over a short time interval under accelerated conditions, such as high ambient temperatures? Fig. 15 shows data obtained on four different suppliers' product when operated at a $+80^{\circ}\text{C}$ ambient temperature. The temperature of the selenium cells under these conditions (ambient plus temperature rise) is about 100°C . For comparison purposes, data are included for similar stacks aged at normal

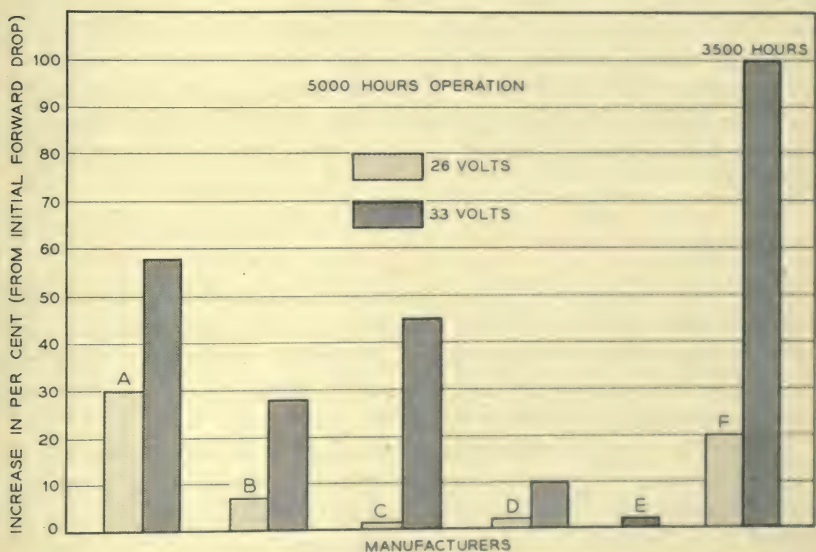


Fig. 12 — Forward aging characteristics. Comparison of 26- and 33-volt cell ratings.

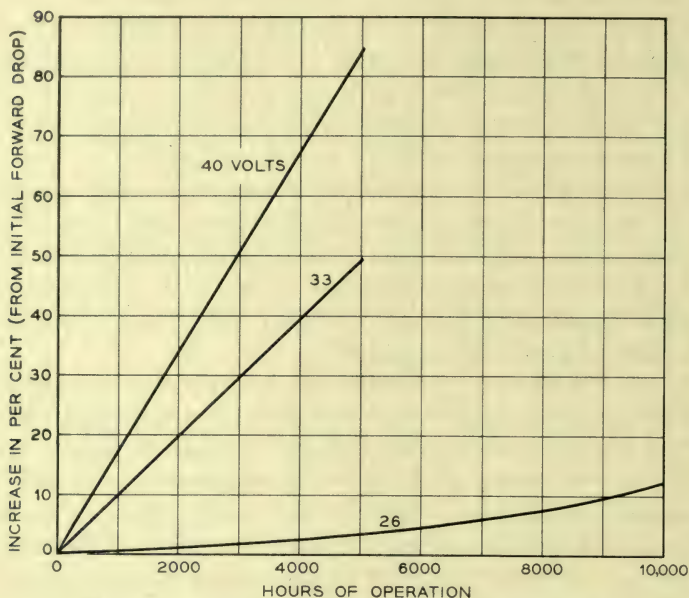


Fig. 13 — Forward aging versus cell voltage rating.

room temperatures. It is readily evident that at $+80^{\circ}\text{C}$ ambient (1) the forward aging rate is increased (2) the different manufacturers' stacks age at different rates and (3) there is no definite correlation of the aging rate with normal operation at room temperature. As previously mentioned under room temperature operation, longer life can be expected if stacks are operated at reduced current loads. At 80°C ambient however, three of the four manufacturers' stacks aged as much at half load as at full load.

CONCLUSIONS

The development and use of selenium rectifier stacks in this country has been quite rapid in the last ten years. The voltage ratings per cell have increased from 14 volts rms to 26 and 33 volts on a commercial basis. The trend in the selenium industry seems to indicate that cell voltage ratings will be further increased.

The proper selection and procurement of selenium rectifier stacks and their application to dc power supplies for large scale industrial or military projects usually require considerably more information than can be obtained from published data in manufacturers' catalogs. Considerable differences have been observed in the performance of similar stacks

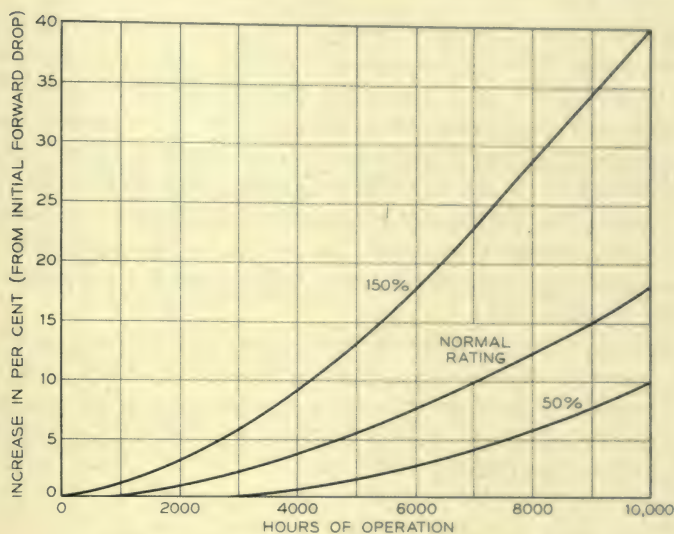


Fig. 14 — Effect of load current on forward aging.

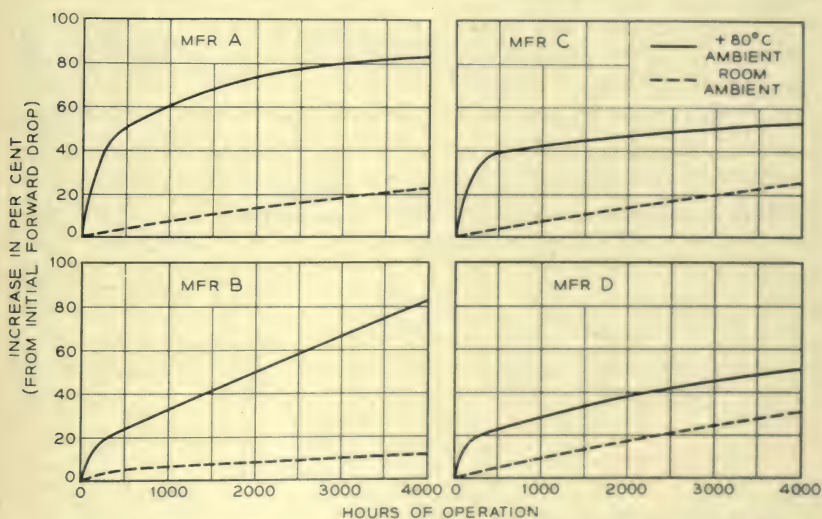


Fig. 15 — Forward aging characteristics. Comparison of room and +80°C ambient temperature operation. 26-volt cells, resistance load, rated dc amps, input voltage — 80 per cent normal for +80°C ambient and 90 per cent normal for room ambient.

furnished by different manufacturers. To appraise properly the qualities of different manufacturers' stacks, the *stability of the life characteristics* should be considered more important than the initial characteristics.

There is very little information published by the selenium manufacturers regarding the life expectancy of their product. Aging appears to be directly related to the individual manufacturing techniques used by each supplier. The life of selenium rectifier stacks seem to decrease as the cell voltage rating is increased. Longer life can be expected if stacks are operated at load currents below the present ratings given in the manufacturers' literature.

Selenium rectifier stacks properly designed and conservatively rated can be expected to give satisfactory performance for 10 to 20 years. Careful consideration of the rate and extent of aging must be evaluated so that proper allowance may be made in the circuit design to obtain maximum life expectancy.

REFERENCES

- E. A. Harty, Characteristics and Applications of Selenium Rectifier Cells, Elec. Eng., Oct., 1943.
- J. H. Hall, Transformer Calculations for Selenium Rectifier Applications, Elec. Eng., Feb., 1946.
- Glen Ramsey, The Selenium Rectifier, Elec. Eng., Dec., 1944.
- J. Gramels, Problems to Consider in Approving Selenium Rectifiers, A.I.E.E. Technical Paper No. 53-215. Also published in Communications and Electronics, Sept., 1953.
- W. F. Bonner, Advanced Developments in Metallic Rectifiers, Electrical Manufacturing, Oct., 1951.
- I. T. Cataldo, Development of 40-volt Selenium Rectifier Plates, Electrical Manufacturing, May, 1952.
- S. Niciejewski, Selenium Rectifiers, Radio and Television News, Oct., 1952.
- H. K. Henisch, Metal Rectifiers, Oxford, Clarendon Press, 1949.
- E. A. Richards, The Characteristics and Applications of the Selenium Rectifier, J. Inst. Elec. Engrs. (London), **88**, Part III, Dec., 1941.

Arcing of Electrical Contacts in Telephone Switching Circuits

Part II—Characteristics of the Short Arc

By M. M. ATALLA

(Manuscript received May 27, 1953)

Results are presented of an experimental study of the characteristics of the short arc in air which is the major cause of contact erosion in telephone switching circuits. Measurements were made of the arc initiation voltage, the voltage drop across the arc and the minimum arcing current. The following are the main conclusions: (1) For "normal" contacts in air, the arc is initiated at a constant field strength of a few million volts/cm up to separations of about 2–3 mean free paths of an electron in air. At larger separations the arc is initiated at the well known spark breakdown potentials of air. In vacuum the linear relation holds for larger separations followed by a transition into a square root relation $V_{ai} = K(d)^{1/2}$. (2) For "clean" contacts in air, no constant field strength line is obtained for separations as low as 1600 Å. Instead, the arc is initiated at the spark breakdown potentials of air, possibly due to adsorbed air molecules or due to breakdown along a longer path at the Paschen's minimum potential. In vacuum, it is speculated that the above square root relation will hold. (3) For "activated" contacts and small separations the arc is initiated at a constant field strength of about 0.6×10^6 volts/cm. (4) For "normal" contacts the minimum arcing current increases with an increase in the maximum current during the arc due to surface contaminations and the arc cleaning action. (5) For arc currents above 1.5 ampere and energies of the order of thousands of ergs the cathode determines the arc characteristics.

INTRODUCTION

The electrical erosion of contacts presents an important problem in the design of telephone switching apparatus. There are several physical phenomena that occur between contacts and contribute to their erosion. The short arc,* which may occur on both make and break of a contact,

* The short arc is characterized by its constant voltage, independent of the current, which is of the order of the ionizing potential of the contact material.

is generally considered to be the major contributor. For illustration, a palladium contact, 10^{-4} cm³ in volume, will last for more than 10^9 operations* only if the arc energy per operation is less than 2.5 ergs. This is based on an erosion rate of 4×10^{-14} cm³ per erg.¹ Furthermore, a short arc with a half ampere current, lasting for only one microsecond, will dissipate as much energy as 70 ergs. Contact erosion may also take place, though at much lower rates for the usual ranges of current and voltage in switching circuits, due to molten bridges² on contact break and due to glow discharge.³

In Part I of this series,⁴ was discussed the mechanism of the initiation of the short arc as determined by contact and circuit conditions. Three characteristics of the arc were used in the presentation without elaboration as to their nature: (1) the arc initiation voltage, (2) the voltage drop across the arc, and (3) the arc initiation and the arc termination currents. These characteristics have been the subject of a recent study to which this part of the series is mainly devoted.

In the course of this study, it was found that there should be some repetition of previous work to isolate effects of certain pertinent parameters that were not previously given due consideration.

No attempt is made here to give a complete survey of the related studies in the literature. Only a few publications are referred to as typical references to the subjects discussed.

NOTATION

C	Capacitance
E_a	Energy dissipated in the arc
F	Gross field strength between the contacts: $\frac{V}{d}$
I	Current
I_i	Arc initiation current
I_{max}	Maximum current in the arc
I_m	Minimum arcing current or arc termination current

* This is the actual life requirement of some contacts in existing switching circuits.

¹ L. H. Germer and F. E. Haworth, Erosion of Electrical Contacts on Make, *J. App. Phys.* **20**, p. 1085, 1949.

² See for example: J. J. Lander and L. H. Germer, The Bridge Erosion of Electrical Contacts, *J. App. Phys.* **19**, p. 910, 1948.

³ F. E. Haworth, Electrode Reactions in Glow Discharge, *J. App. Phys.* **22**, p. 606, 1951.

⁴ M. M. Atalla, "Arching of Electrical Contacts in Telephone Switching Circuits. Part I—Theory of the Initiation of the Short Arc," *B.S.T.J.*, **32**, pp. 1231–1244, Sept., 1953.

K	Constant in the relation $V_{ai} = K(d)^{1/2}$
R	Resistance
V	Voltage
V_{ai}	Arc initiation voltage
d	Minimum separation between contacts
h	Height of a metal bridge formed during one arc
t	Time
t_a	Arc duration
v	Constant voltage drop across the short arc

ARC INITIATION VOLTAGE

Consider the simple contact circuit in Fig. 1 comprising a pair of contacts in series with a resistor R and a variable voltage power supply. By fixing the separation between the contacts and gradually increasing the

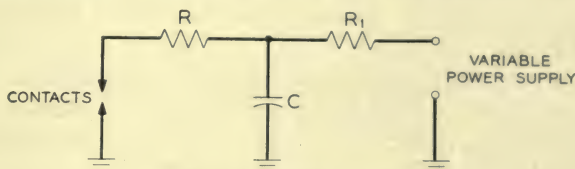


Fig. 1 — Contact Circuit.

voltage, an arc is usually initiated when the voltage reaches a certain value " V_{ai} ," called the arc initiation voltage. In general V_{ai} is a function of: (1) separation between the contacts, (2) geometry of the contact surfaces, (3) the surrounding atmosphere, and (4) contact material and its surface.

Our experiments were limited to contacts in atmospheric air with special emphasis on separations of the order of and less than the mean free path of an electron in air. For larger separations the arc initiation voltage follows the well known curve of the sparking potential of air.⁵ For the smaller separations where the presence of air molecules would not be expected to affect the arc initiation voltage, it was previously reported that breakdowns occurred at some constant field between 0.6×10^6 and 16×10^6 volts/cm.^{1, 6, 7}

In our experiments a cantilever bar, described by Pearson,⁶ was used. The setting for zero separation was determined by a 0.1 volt source, a

⁵ See for instance J. D. Cobine, *Gaseous Conductors*, McGraw-Hill, N. Y., p. 162, 1941.

⁶ G. L. Pearson, *Phys. Rev.* **56**, p. 471, 1939.

⁷ L. H. Germer and J. L. Smith, *J. App. Phys.* **23**, p. 553, 1952.

10,000 ohms resistor and a cathode relay oscilloscope. The zero setting could be repeated with a precision of $\pm 500\text{\AA}$. All the reported results were obtained by fixing the contact separation, raising the voltage and observing the breakdown on a cathode ray oscilloscope. Contacts tested were given one of three different surface treatments:

(1) The contact surface was polished with fine emery paper, washed with methyl alcohol, then exposed to the laboratory atmosphere for a few hours. After about 5 arc discharges, readings of arc initiation voltage seemed to vary at random. Contacts thus treated are referred to as "normal" contacts.

(2) Contacts were subjected to arcing for about 5 minutes at the rate of 15 arcs per second. Arcing was produced by the discharge of a half microfarad condenser at 500 volts through a 10-ohm resistor. Measurements of the arc initiation voltage followed *immediately after* this treatment. These contacts are referred to as "clean" contacts. Their behaviour usually changed to that of "normal" contacts after a short exposure to the laboratory atmosphere.

(3) Contacts were subjected to arcing at the rate of 3 arcs per second for about one hour in air saturated with d-limonene. The arc was produced by discharging a 0.1-microfarad condenser at 50 volts through a 100-ohm resistor. These contacts are referred to as "activated" contacts.⁸

Fig. 2 shows the results obtained with "normal" palladium contacts. Each point represents the average of five readings. The maximum spread was 40 per cent of the average. For separations less than $10,000\text{\AA}$, about two mean free paths of an electron in air at normal conditions, a constant gross field strength line of 3×10^6 volts/cm was obtained. At larger separations the measured arc initiation voltages were essentially the well known sparking potentials of air. Fig. 3 shows the corresponding results obtained with "normal" carbon contacts in air. Below a separation of $15,000\text{\AA}$, the arc was initiated at a constant gross field strength of 2.4×10^6 volts/cm. The maximum spread of the individual points was only 15 per cent of the average.

In the absence of air, it is expected that the constant field strength lines will hold for higher separations.* In Table I, Column 2 are given the measured values of the gross field strengths at which the arc was initiated for a group of "normal" contact materials.

* Recent unpublished measurements by Dr. P. Kisliuk on similar contacts in vacuum have indicated that the constant field strength relation $V_{ai} = F(d)$ holds initially for larger separations and is followed by a gradual transition into a square root relation $V_{ai} = K(d)^{1/2}$ as proposed by Cranberg.⁹

⁸ L. H. Germer, Arcing at Electrical Contacts on Closure—Part I, J. Appl. Phys. **22**, p. 955, 1951.

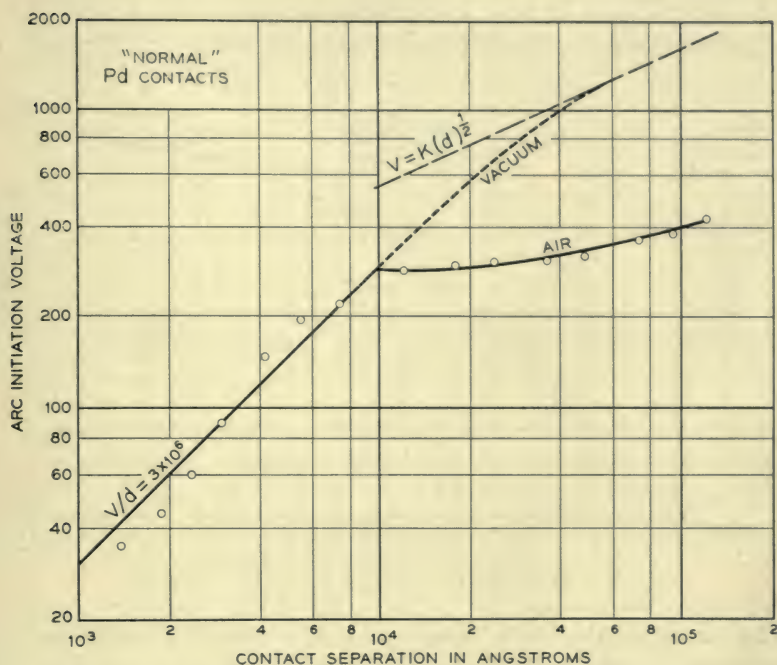


Fig. 2 — Arc initiation voltages for "Normal" palladium contacts.

Due to the spread of the above measurements and their observed dependence on surface exposure and treatment it was suspected that the constant field strength characteristic obtained was due to surface contamination. This was verified by testing contacts cleaned by the heavy arcing process explained above. The results are shown in Fig. 4. The familiar constant field strength line *was not* obtained for separations as low as $1500\text{\AA}$. Instead, the arc was initiated at voltages comparable to the sparking potentials of air. Since the separations were too small, the smallest being three times less than the mean free path of an electron in air, it was thought that the effect was due to some mechanism involving the adsorbed air molecules or due to breakdown along a longer path at the Paschen's minimum potential.⁵ In the absence of air, it is, therefore, expected that higher voltages and higher field strengths in excess of 20×10^6 volts/cm, as obtained at $1500\text{\AA}$, will be needed to initiate the arc. It is possible that Cranberg's relation⁹ $V = K(d)^{1/2}$, will hold for separations as low as a few thousand angstroms. In Fig. 4, this relation, with $K = 10^5$ volt. cm^{-1/2}, is plotted.

⁹ L. Cranberg, The Initiation of Electrical Breakdowns in Vacuum, J. Appl. Phys. **23**, p. 518, 1952.

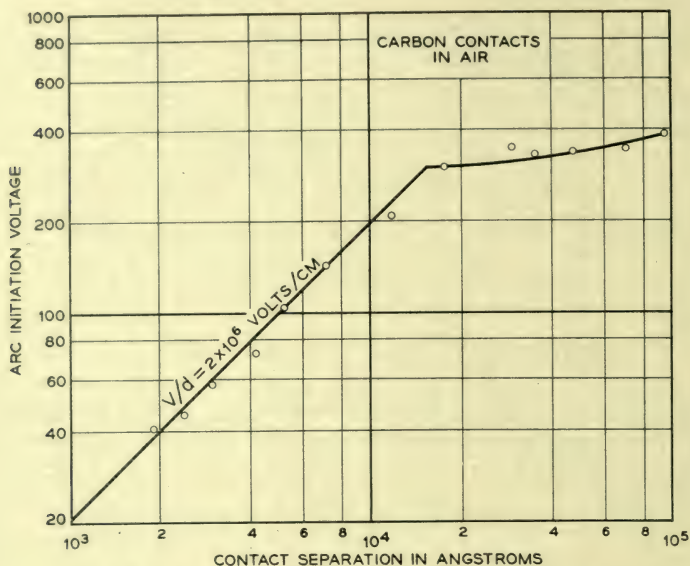


Fig. 3 — Arc initiation voltages for carbon contacts.

For contacts activated in organic vapors, constant field strength lines were obtained. In Fig. 5 results are shown for palladium contacts activated by d-limonene. The average field strength for arc initiation is only 0.6×10^6 volts/cm with a spread of as much as 100 per cent of the average. At separations greater than 50,000 Å the arc was initiated by the familiar spark breakdown of air. In vacuum the constant field strength line should hold for larger separations possibly until it intersects Cranberg's line.⁹

Breakdowns at low fields were also observed for metals with inorganic films. For instance Gleichauf¹⁰ obtained a constant field strength line of 0.24×10^6 volts/cm for copper electrodes in vacuum at separations of the order of millimeters. Our measurements on copper contacts at separations less than 10,000 Å have shown breakdowns at fields as low as 0.7×10^6 volts/cm. It is concluded that the presence of organic or inorganic films on a contact usually leads to a reduced gross field strength at which the breakdown will occur. The reduction can be by as much as two orders of magnitude. It is possible that this reduction was only apparent and the electrons actually came from the underlying metal by extraction in an intense field set up by the positive ions lying on the surface of the film.¹¹

¹⁰ P. H. Gleichauf, Electrical Breakdown Over Insulators in High Vacuum, J. Appl. Phys. **22**, p. 766, 1951.

¹¹ F. L. Jones, Electrical Discharges in Gases, Nature, **170**, p. 601, 1952.

TABLE I. — ARCING CHARACTERISTICS OF CONTACT MATERIALS

(1) Contact Material	(2) Field strength to initiate the arc for normal contacts. 10 ⁶ Volts/cm.	(3) Short arc Voltage	(4) Minimum arcing current for "clean"* contacts. Amps.
Carbon.....	2.4	20-43	0.03
Nickel.....	4.2	12-13	0.4
Palladium.....	3.0	14-15	1.1
Silver.....	2.0	11-13	0.8
Tungsten.....	4.9	12-13	0.7

* For "normal" contacts, the minimum arcing current is less by 50 per cent or more.

The results of the above section are summarized in Fig. 6. The solid lines were actually measured for palladium contacts under different surface conditions. The broken lines are only speculative. For a certain separation and surface condition the arc will be initiated at a voltage as given by the lowest corresponding line in the figure.

VOLTAGE DROP ACROSS A SHORT ARC

The short arc may be defined as a discharge of electricity between electrodes with a voltage drop of the order of the minimum ionizing potential of the atoms of the electrodes.* Furthermore, due to the small separation between the contacts and the local high pressure metal vapor, the characteristics of the established arc are independent of a surrounding atmosphere at normal or low pressures. The short arc is characterized by its constant voltage for currents above a minimum value called the minimum arcing current of the contact. In contrast to the short arc, the long arc between contacts at a fixed separation has a voltage drop which decreases with an increase in current.¹² *Most arcs occurring between contacts on both make and break of telephone switching circuits, are short arcs.* In Table I, Column 3, are given our measured values of the short arc voltage for a few materials.

ARC INITIATION AND TERMINATION CURRENTS

The arc termination current or minimum arcing current is defined as the lowest current at which the arc can be sustained. The arc is extin-

* The arc voltage is about 50 per cent higher except for the carbon arc which has a much higher voltage; see Table I, Column 3.

¹² See for instance: K. Gaulrapp, *Untersuchung der elektrischen Eigenschaften des Abreisbogens*, Ann. Physik. **25**, p. 705, 1936.

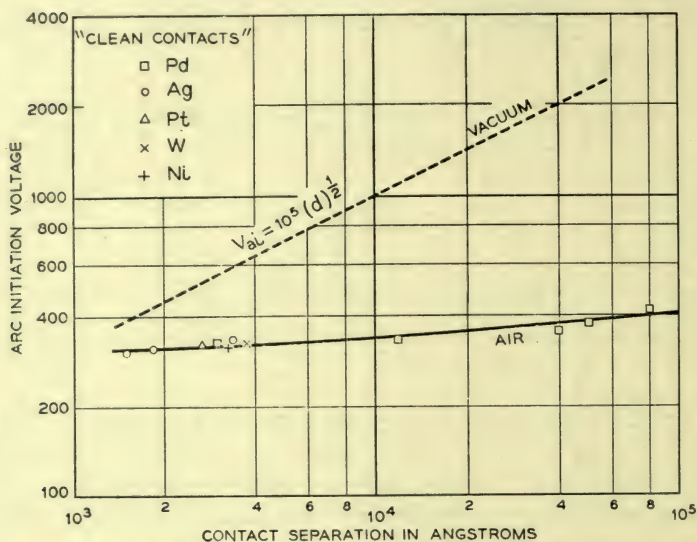


Fig. 4 — Arc initiation voltages for "Clean" metals.

guished when the circuit current drops to this value. To initiate the arc a minimum current must be furnished by the circuit called the arc initiation current. It was previously shown⁴ that the arc initiation and termination currents are essentially the same numerically. The existence of these limiting currents as such, rather than current densities, is not understood from a fundamental standpoint.

The limiting currents are a function of the contact material and are appreciably affected by the surface condition of the contacts. Surface contaminations generally reduce the limiting currents of the contacts. The results presented here were obtained by measuring the residual voltage in an R-C circuit following an arc. This voltage is equal to $I_m R$, from which I_m was determined. Our measurements are given in Table I, Column 4 for "clean" contacts. The maximum spread is 20 per cent of the average. For normal contacts, however, surface contamination causes a wide variation in the results. Furthermore, the maximum intensity of the arc, or the maximum current furnished by the measuring circuit, was found to have an appreciable affect on the measurements. This appears to be due to the surface cleaning action of the arc. This effect is demonstrated in Fig. 7 for "normal" palladium contacts. Each point represents an individual measurement of I_m plotted against the corresponding maximum current during the arc. An R-C contact circuit was used. While the measurements show a considerable spread, they indicate a definite trend of an increase in I_m with increasing I_{max} .

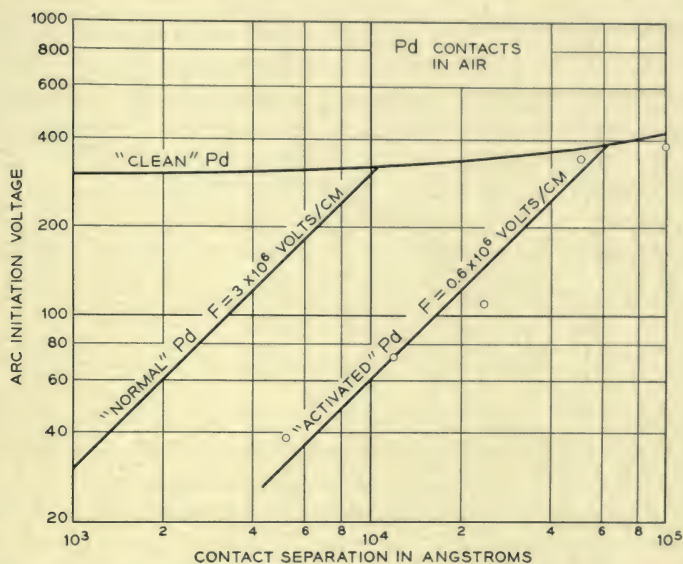


Fig. 5 — Arc initiation voltages for "Activated" palladium contacts.

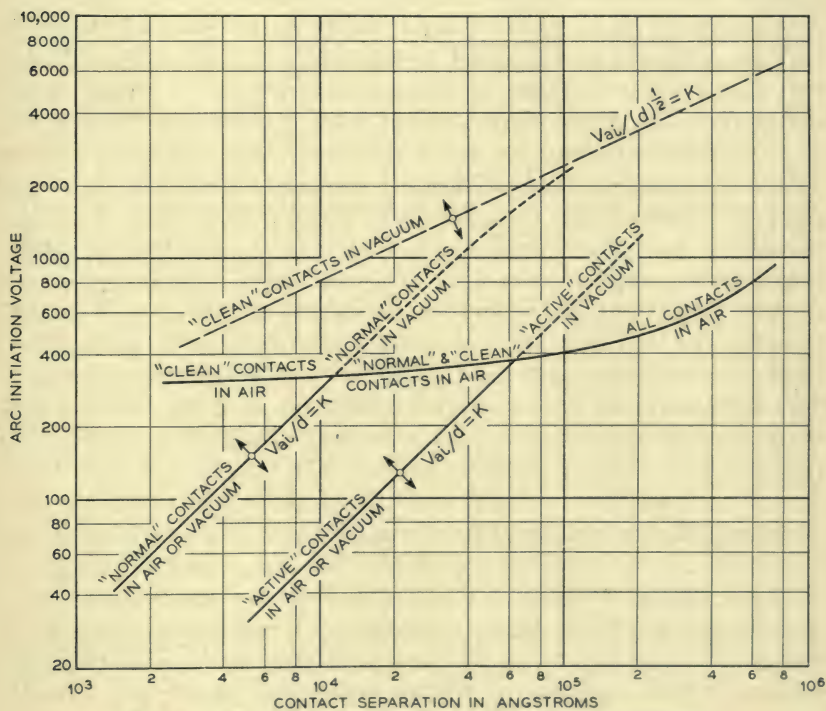


Fig. 6 — Summary of results of arc initiation voltages.

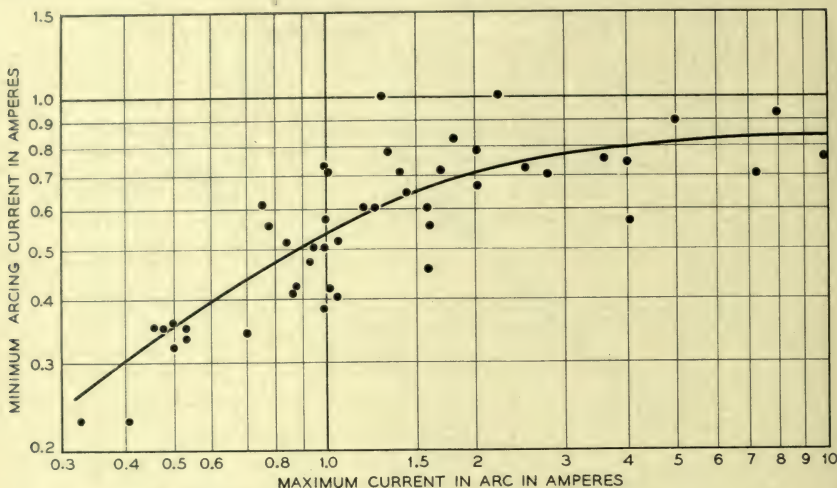


Fig. 7 — Dependence of minimum arcing current on maximum arc current.

In the first part of this paper it was shown that contact activation by organic vapors reduces the arc initiation voltage for a fixed separation. In other words, for a pair of closing contacts the arc will be initiated at a wider separation and a longer arcing time is obtained. In addition, activation tends to decrease the minimum arcing current. Germer⁸, measured a minimum arcing current of only 0.027 to 0.037 ampere for active silver. Our measurements for active palladium gave a minimum arcing current of 0.1 ampere. This substantial decrease in the minimum arcing current of contacts due to surface activation usually causes a further increase in the arcing time. Contact activation, therefore, enhances arcing between closing contacts in two ways; first, by initiating the arc at wider separations, and second by maintaining the arc at much smaller currents. The following results are presented to indicate the quantitative significance of contact activation. A pair of "normal" palladium contacts were operated in air saturated with d-limonene at 3 cps. The contacts closed a circuit consisting of a 0.5-microfarad condenser, charged to 50 volts, in series with a 100-ohm resistor. The transient on make was observed on a cathode ray oscilloscope to determine the arcing time. The arc energy E_a was calculated and the ratio $E_a/Cv (V_0 - v)$ was plotted against the number of operations, Fig. 8. The denominator is the maximum arc energy, which is only attained if the arc is maintained until the current reaches zero. The results indicate a rapid increase of the arc energy corresponding to an increase in surface activation. When the contacts become fully active, the arc energy was about two orders of

magnitude greater than the energy for inactive contacts. Contacts may also be activated by inorganic films. Experiments on palladium and silver contacts have shown* that glow discharge between contacts operating in air produced a second type of activation. Nitrides were formed on the contact surfaces and the minimum arcing current dropped to about 0.1 ampere for silver and 0.2 ampere for palladium. This effect was more pronounced with silver contacts.

In carrying out the measurements of the minimum arcing current it was observed that the arc was generally interrupted by one of three causes (1) the minimum arcing current was reached, (2) physical closure of the contacts, and (3) shorting of the arc by a metal bridge formed during the arc. A study of the bridge formation has shown that the height of the bridge was a function of the arc energy. The bridge height was measured by setting the zero separation point before and after the arc. The difference gave the height of the bridge. This was plotted in Fig. 9 against the measured arc energy. The height of the bridge increased roughly with the cubic root of the arc energy up to energies of

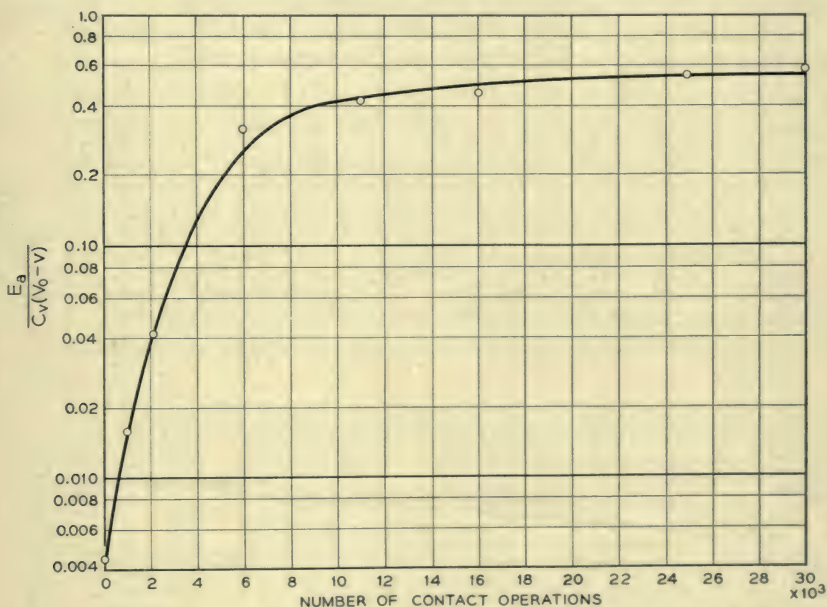


Fig. 8 — Increase in arc energy due to contact "Activation".

* This information was obtained from unpublished work of F. E. Haworth. See also Reference 8 for a discussion of the effects of insulating films on arc initiation.

about 800 ergs. This was followed by a rapid transition into metal *loss* instead of a bridge. This loss increased with increase in energy.

DISSIMILAR CONTACTS — EFFECT OF POLARITY ON ARC CHARACTERISTICS

An experiment was carried out to find the contributions of the anode and cathode in determining the characteristics of the arc. Use was made of the fact that carbon contacts are unique in having low minimum arcing current and high arc voltage compared to most metals; see Table I. Furthermore, carbon contacts always gave a constant field strength line for arc initiation at small separations even when they were cleaned by the heavy arcing process. Tests were carried out with a variety of contact metals against carbon. All the results obtained were qualitatively the same. For illustration only the experiments with palladium-carbon contacts are reported here.

Test specimens were carefully prepared in the following fashion. Palladium to palladium contacts, mounted on the cantilever bar set-up, were cleaned by the heavy arcing process. One contact was removed and replaced by a carbon contact with its surface polished and freed from

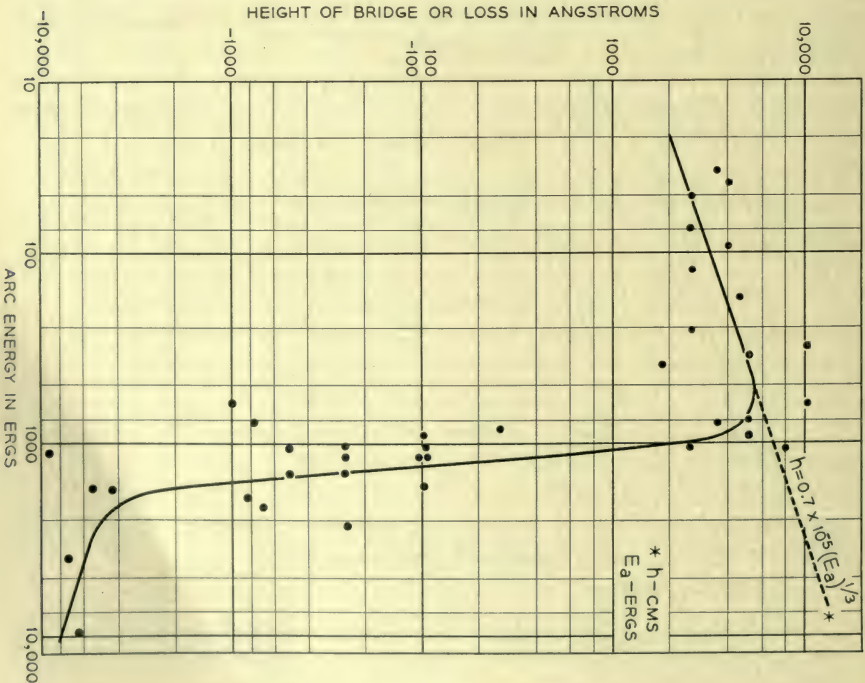


Fig. 9 — Bridge formation during the arc.

TABLE II — EFFECT OF POLARITY ON THE CHARACTERISTICS OF THE ARC — PALLADIUM-CARBON CONTACTS

(1) Contact Configuration	(2) Arc initiation voltage at 6000Å separation	(3) Arc voltage v	Minimum arcing current, I_m amps.
C ⁺ , Pd ⁻	300	13-15	0.2-0.5
Pd ⁺ , Pd ⁻	320	14-15	1.1
C ⁻ , Pd ⁺	130	20-30	0.2-0.3
C ⁻ , C ⁺	120	20-43	0.03

loose particles. The separation between the contacts was set at 6000Å. By gradually raising the voltage across the contacts and observing it on a cathode ray oscilloscope, the arc initiation voltage was determined. The same measurement was repeated several times. Preceding *each* measurement, the contacts were recleaned by the same process explained above. The polarity was then reversed and a new set of measurements was taken. The results are shown in Table II, Column 2. They indicate that the arc initiation voltage is determined by the cathode. This furnishes support to the postulate that field emission is the first step of the mechanism of arc initiation. By recording the voltage across the contacts during the arc, measurements were made of the arc voltage and the minimum arcing current. The results are given in Table II, Columns 3 and 4. The arc voltage measurements indicate rather conclusively that the arc voltage is determined by the cathode. The minimum arcing current measurements, however, were only slightly, yet consistently, higher with a palladium cathode. It is thought that during a single short arc, particularly with the high intensity arcs used, there is a certain amount of exchange of materials between the electrodes. This exchange is possibly responsible for the observed influence of the anode on the arc characteristics.

It is concluded that the arc initiation voltage as well as the arc voltage are characteristics of the cathode while the minimum arcing current seems to be influenced by both electrodes with stronger inclination towards the cathode characteristics.* The following reservation, how-

* Early experiments by H. E. Ives¹³ have led to the conclusion that the arc voltage is a characteristic of the *anode*. In his experiments, however, currents of the order of one ampere were established in the circuit while the *contacts were closed*. Arc measurements were made during the subsequent break of the contacts. It is thought, therefore, that metal bridges must have formed during the break, transferring metal from the anode to the cathode.² The arc produced must have been influenced accordingly. In our experiments this difficulty was entirely eliminated.

¹³ H. E. Ives, Minimal Length Arc Characteristics, J. Franklin Inst. **198**, No. 4, 1924.

ever, has to be made. All these measurements, although made at a voltage range between 50 and 400 volts corresponding to a contact separation range between 1500 and 100,000Å, had high maximum currents above 1.5 amperes. The energy dissipated in each arc was of the order of thousands of ergs. This reservation is made because recent studies of metal transfer have indicated a reversal in the direction of transfer between the anode and the cathode depending on the rate of energy dissipation in the arc.

ACKNOWLEDGEMENTS

I am indebted to L. H. Germer, P. Kisliuk and F. H. Haworth for much valuable discussion and to A. S. Timms for assistance with many of the experiments reported here.

Abstracts of Bell System Technical Papers*

Not Published in this Journal

ANDERSON, J. R.¹

Electrical Delay Lines for Digital Computer Applications, I.R.E., Trans. P.G.E.C., **2**, pp. 5-13, June, 1953.

A survey of existing lumped parameter and distributed parameter delay lines has shown that their maximum storage capacity is about 23 pulses and 15 pulses respectively regardless of total delay time. An analysis of pulse transmission through distributed delay lines indicates that dissipation in the inductive elements is the chief factor limiting storage capacity. A method is proposed for decreasing this dissipation through the use of high Q nickel zinc ferrites around straight conductors for inductive elements.

ARMSTRONG, C. A.²

Communications for Civil Defense, A.I.E.E., Trans., Commun. & Electronics, **7**, pp. 315-326, July, 1953.

BAROTTA, P. J., see S. P. GENTILE.

BIRDSALL, H. A., see D. A. McLEAN.

BOGERT, B. P.¹

On the Band Width of Vowel Formants, Letter to the Editor, J. Acoust. Soc. Am., **25**, pp. 791-792, July, 1953.

Measurements were made on a sample of vowel utterances, by male talkers, of the band widths of the first three formants. It was found that the band

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

¹ Bell Telephone Laboratories.

² American Telephone and Telegraph Company.

width was essentially constant and independent of the particular vowel. The mean values for bars 1, 2, and 3 were 130, 100, and 185 cps, respectively. Ten per cent of the 300 band widths measured were less than 90 cps and ten per cent greater than 260 cps.

BRANGACCIO, D. J., see C. C. CUTLER.

BRATTAIN, W. H., see A. M. PORTIS.

BULLINGTON, K.¹

Frequency Economy in Mobile Radio Bands, I.R.E., Trans., P.G.V.C., **3**, pp. 4-27, June, 1953 (Monograph 2109).

CALBICK, C. J., see D. A. McLEAN.

CAMPBELL, R. D.²

Path Testing for Microwave Radio Routes, Elec. Eng., **72**, pp. 571-577, July, 1953.

CAMPBELL, W. E.¹

Solid Lubricants, Lubrication Eng., **9**, pp. 195-200, Aug., 1953.

CICCOLELLA, D. F.¹ AND L. J. LABRIE¹

High Frequency Crystal Units for Use in Selective Networks and Their Proposed Application in Filters Suitable for Mobile Radio Channel Selection, I.R.E., Trans, P.G.V.C. **3**, pp. 118-128, June, 1953.

CONWELL, E. M.⁶

High Field Mobility in Germanium With Impurity Scattering Dominant, Phys. Rev., **90**, pp. 769-772, June 1, 1953.

Experimental measurements show a variation of mobility with electric field intensity of electrons in *n* type germanium which differs at 20 degrees K from that observed in the same specimen at 77 degrees K and higher temperatures. This difference can be accounted for by scattering by ionized impurities. A crude quantitative treatment is carried out along the lines of Shockley's treatment for the case of lattice scattering. As in that case, the resulting theory fits the data well if the rate of energy loss is taken several times higher than that given by the theory assuming that the surfaces of constant energy are spherical.

¹ Bell Telephone Laboratories, Inc.

² American Telephone and Telegraph Company.

⁶ Formerly Bell Telephone Laboratories.

CUTLER, C. C.¹ AND D. J. BRANGACCIO¹

Factors Affecting Traveling Wave Tube Power Capacity, I.R.E., Trans., P.G.E.D., **3**, pp. 9-23, June, 1953.

DACEY, G. C.¹

Space-Charge Limited Hole Current in Germanium, Phys. Rev., **90**, pp. 759-763, June 1, 1953.

A situation can arise in semi-conductors similar to the space-charge limited emission of electrons in vacuum. The theory of Shockley and Prim for this phenomenon has been extended to the high field case using the approximation that the drift velocity of the carriers is $v = \mu(EE_0)^{1/2}$, where μ is the low field mobility, E the electric field, and E_0 the "critical field." For this approximation the current density analogous to Child's law for a plane parallel diode is

$$J = (2/3)(5/3)^{3/2} K \mu E_0^{1/2} V_a^{3/2} / w^{5/2},$$

where V_a is the potential across a diode of thickness w and K is the dielectric constant in mks units. Good agreement between theory and experiment for hole flow in germanium at liquid air temperature has been obtained, using values of μ and E_0 obtained independently by Ryder.

ESHELBY, J. D.⁸, READ, W. T.¹ AND W. SHOCKLEY¹

Anisotropic Elasticity with Applications to Dislocation Theory, Acta Metallurgica, **1**, pp. 251-259, May, 1953.

The general solution of the elastic equations for an arbitrary homogeneous anisotropic solid is found for the case where the elastic state is independent of one (say x_3) of the three Cartesian coordinates x_1, x_2, x_3 . Three complex variables $z(\ell) = x_1 + p(\ell)x_2$ ($\ell = 1, 2, 3$) are introduced, the $p(\ell)$ being complex parameters determined by the elastic constants. The components of the displacement (u_1, u_2, u_3) can be expressed as linear combinations of three analytic functions, one of $z_{(1)}$, and of $z_{(2)}$, and one of $z_{(3)}$. The particular form of solution which gives a dislocation along the x_3 -axis with arbitrary Burgers vector (a_1, a_2, a_3) is found. (The solution for a uniform distribution of body force along the x_3 -axis appears as a by-product.) As is well known, for isotropy we have $u_3 = 0$ for an edge dislocation and $u_1 = 0, U_2 = 0$ for a screw dislocation. This is not true in the anisotropic case unless the x_1x_2 plane is a plane of symmetry. Two cases are discussed in detail, a screw dislocation running perpendicular to a symmetry plane of an otherwise arbitrary crystal, and an edge dislocation running parallel to a fourfold axis of a cubic crystal.

FULLER, C. S.¹ AND J. A. DITZENBERGER¹

Diffusion of Lithium Into Germanium and Silicon, Letter to the Editor, Phys. Rev., **91**, p. 193, July 1, 1953.

¹ Bell Telephone Laboratories, Inc.

⁸ University of Illinois.

GENTILE, S. P.¹ AND P. J. BAROTTA⁴

Transistor Physics Simplified, Radio and Telev. News, **50**, pp. 44-46, 100-102, July, 1953.

HINES, M. E.¹

Traveling-Wave Tube, Radio and Telev. News, **49**, pp. 12-14, 26, June, 1953.

KOLB, E. D., see W. P. SLICHTER.

LABRIE, L. J., see D. F. CICCOLELLA.

LINVILL, J. G.¹

Transistor Negative-Impedance Converters, I.R.E., Proc., **41**, pp. 725-729, June, 1953.

MAY, A. S.¹

Microwave System Test Equipment, Commun. Eng., **13**, pp. 24-25, 44-45, May-June, 1953.

McKAY, K. G.¹

Bombardment Conductivity, Ind. Diamond Rev., **13**, pp. 127-130, June, 1953.

McLEAN, D. A.,¹ BIRDSALL, H. A.¹ AND C. J. CALBICK¹

Microstructure of Capacitor Paper, Ind. and Eng. Chem., **45**, pp. 1509-1515, July, 1953 (Monograph 2142).

McMILLAN, B.¹

Basic Theorems of Information Theory, Ann. Math. Stat., **24**, pp. 196-219, June, 1953 (Monograph 2124).

This paper describes briefly the current mathematical models upon which communication theory is based, and presents in some detail an exposition and partial critique of C. E. Shannon's treatment of one such model. It then presents a general limit theorem in the theory of discrete stochastic processes, suggested by a result of Shannon's.

¹ Bell Telephone Laboratories, Inc.

⁴ Hudson Technical Institute.

MERTZ, P.¹

Influence of Echoes on Television Transmission, J.S.M.P.T.E., 60, pp. 572-596, May, 1953 (Monograph 2144).

PETERSON, G. E.¹

Basic Physical Systems for Communication Between Two Individuals, J. Speech and Hearing Disorders, 18, pp. 116-120, June, 1953 (Monograph 2135).

PIERCE, J. R.¹

Transistors, Radio-Electronics, 24, pp. 42-44, June, 1953.

PORTIS, A. M.,⁷ A. F. KIP,⁷ C. KITTEL,⁷ AND W. H. BRATTAIN¹

Electron Spin Resonance in a Silicon Semi-Conductor, Letter to the Editor, Phys. Rev., 90, pp. 988-989, June 1, 1953.

PRIM, R. C., see W. SHOCKLEY.

QUARLES, D. A.⁵

Progress and Problems, Elec. Eng., 72, pp. 667-669, August, 1953.

QUARLES, D. A.⁵

Report to the Membership, Elec. Eng., 72, pp. 477-479, June, 1953.

READ, W. T., see J. D. ESHELBY.

RYDER, E. J.¹

Mobility of Holes and Electrons in High Electric Fields, Phys. Rev., 90, pp. 766-769, June 1, 1953.

The field dependence of mobility has been determined for electrons and holes in both germanium and silicon. The observed critical field at 298 degrees K beyond which μ varies as $E^{-1/2}$ is 900 volts/cm for *n*-type germanium, 1400 volts/cm for *p*-type germanium, 2500 volts/cm for *n*-type silicon, and 7500 volts/cm for *p*-type silicon. These values of critical field are between two to four times those calculated on the basis of spherical constant energy surfaces in the Brillouin zone. A saturation drift velocity of $6(10)^6$ cm/sec is observed in germanium which is in good agreement with predictions based on scattering

¹ Bell Telephone Laboratories, Inc.

⁵ Sandia Corporation.

⁷ University of California, Berkeley.

by the optical modes. Data on n -type germanium at 20 degrees K show a range over which impurity scattering decreases and the mobility increases with field until lattice scattering dominates as at the higher temperatures.

SHOCKLEY, W., see J. D. ESHELBY.

SHOCKLEY, W.¹ AND R. C. PRIM¹

Space-Charge Limited Emission in Semi-Conductors, Phys. Rev., **90**, pp. 753-758, June 1, 1953.

A situation analogous to thermionic emission into vacuum can occur in semi-conductors. A semi-conductor analog for a plane parallel vacuum diode may consist of two layers of n type semi-conductor bounding a plane parallel slab of pure semi-conductor. The current density analogous to Child's law is $J = 9\kappa\epsilon_0\mu V^2/8W^3$, where κ = dielectric constant, ϵ_0 = mks permittivity, μ = mobility, V = applied voltage, and W = thickness of pure region. The condition prevailing at the space-charge maximum is analyzed taking into account diffusion due to random thermal motion. Brief discussions are given of the effect of fixed space charge, the dependence of mobility upon electric field strength and the role of space-charge limited emission in a new class of unipolar transistors.

SLICHTER, W. P.¹ AND E. D. KOLB¹

Solute Distribution in Germanium Crystals, Letter to the Editor, Phys. Rev., **90**, pp. 987-988, June 1, 1953.

TOWNSEND, J. R.¹

A Dynamic Program for Conversion, Metal Progress, **63**, pp. 79-81, June, 1953.

TURNER, E. H.¹

New Non-Reciprocal Waveguide Medium Using Ferrites, Letter to the Editor, I.R.E., Proc., **41**, p. 937, July, 1953.

WANNIER, G. H.¹

Threshold Law for Single Ionization of Atoms or Ions by Electrons, Phys. Rev., **90**, pp. 817-825, June 1, 1953.

When an electron hits an atom or ion, it may knock off an electron. This process is fundamental in almost all types of gas discharge. The reaction is endothermic; hence there is a threshold value in the electron energy below

¹ Bell Telephone Laboratories, Inc.

which it does not occur. In this paper, the dependence of the yield on the energy just above this threshold is derived. The derivation is not rigorous because it circumvents some of the difficulties of the three-body problem by applying ergodicity, albeit in a weakened form. The result is that, for atoms, the yield rises as the 1.127th power of the energy excess. For ions the exponent lies between this number and unity.

WILLARD, G. W.¹

Ultrasonically Induced Cavitation in Water: A Step-by-Step Process, J. Acoust. Soc. Am., **25**, pp. 669-686, July, 1953.

A 2.5-mc, barium-titanate, spherically focusing radiator was used to produce cavitation in both degassed and aerated water entirely within the restricted, high intensity focal region, remote from the water boundaries. The sonic intensity rises to 1.8 kw/cm² and the pressure amplitude to ± 70 atmospheres at the focus. High-intensity illumination and an unusual high speed photographic technique permit observation and timing of the step-by-step process of cavitation development.

Feather-shaped cavitation bursts are sporadically produced, being initiated in the insignificant quill portion nearest the radiator, then abruptly expanding to form the catastrophic plume portion. The plume is believed to be formed by myriads of microcavities, too small and close for individual observation. These two fundamental steps are identically produced, and with equal ease, both in degassed and aerated water. The whole action is over in several milliseconds, except that in the case of aerated water a third bubble step is produced. In aerated water, non-collapsing gas bubbles are generated by and concurrently with, the catastrophic step. These bubbles remain after collapse of the burst, to be blown off down stream by the sonically induced liquid streaming.

The bubble step is not generated without the presence of the catastrophic step. The latter is generated only if the initiation step reaches a definite degree of development (not always attained). This requires sonic activation for increasing lengths of time for decreasingly smaller sonic intensities. Origination of the initiation step, and hence of the whole cavitation phenomena, is believed to occur whenever a stray nucleus (weak spot) streams into the high intensity sonic field.

YOUNG, W. R.¹

Comparison of Mobile Radio Transmission at 150, 450, 900 and 3700 Mc, I.R.E., Trans., P.G.V.C. **3**, pp. 71-83, June, 1953.

¹ Bell Telephone Laboratories, Inc.

Contributors to this Issue

M. M. ATALLA, B.S., Cairo University, 1945; M.S., Purdue University, 1947; Ph.D., Purdue University, 1949; Studies at Purdue undertaken as the result of a scholarship from Cairo University for four years of graduate work. Bell Telephone Laboratories, 1950-. For the past three years he has been a member of the Switching Apparatus Development Department, in which he is supervising a group doing fundamental research work on contact physics and engineering. Current projects include fundamental studies of gas discharge phenomena between contacts, their mechanisms, and their physical effects on contact behavior; also fundamental studies of contact opens and resistance. In 1950, an article by him was awarded first prize in the junior member category of the A.S.M.E. He is a member of Sigma Xi, Sigma Pi Sigma, and Pi Tau Sigma, and a junior member of the A.S.M.E.

FRANK E. BLOUNT, B.S. in E.E., Oregon State College, 1928. Bell Telephone Laboratories 1928-. Mr. Blount tested panel system circuits for a year and then transferred to development work on special switching and signaling circuits. His interest in circuit design has also included circuits for automatic toll ticketing for the step-by-step system, for radar testing, and more recently for No. 5 crossbar system. Member of Tau Beta Pi, Eta Kappa Nu and Phi Kappa Phi.

J. T. LINDSAY BROWN, B.S. College of the City of New York, 1915. Western Electric Company, 1915-1925. Bell Telephone Laboratories 1925-. For twenty-five years Mr. Brown was concerned with the development and testing of telephone instruments and such allied apparatus as loudspeakers, microphones, and earphones. In 1940 he transferred to work on the development of glass sealed magnetic switches. He is currently in charge of a group developing mercury contact relays. Member of the A.I.E.E., I.R.E. and Acoustical Society of America.

WALLACE A. DEPP, B.S. and M.S. in E.E., University of Illinois, 1936 and 1937. Bell Telephone Laboratories, 1937-. In his early laboratories

association Mr. Depp was concerned with thoriated tungsten and tantalum emitters and later with cold cathode tubes. During World War II he worked on pulsing thyratrons and fixed spark gap tubes for radar, and miniature thyratrons used in the proximity fuse. He was subsequently in charge of the basic development of all types of gas-filled tubes. Recently he transferred to Transmission Systems Development with responsibility for broad band carrier terminal equipment, N and O carrier systems and automatic switching for the L3 coaxial cable system. Member of the A.I.E.E., Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi and Sigma Xi. Senior member of the I.R.E.

JAMES M. EARLY, B.S. cum laude, New York State College of Forestry, 1943; M.S. and Ph.D. Ohio State University, 1948 and 1951. Bell Telephone Laboratories 1951-. After teaching Electrical Engineering at Ohio State University for five years while studying for his Master's and Ph.D. degrees, Dr. Early joined an electronic apparatus development group, participating in the development of the junction transistor. At present he is doing theoretical as well as development work on high frequency junction transistors. Member of the I.R.E. and Eta Kappa Nu. Associate of Sigma Xi.

JOSEPH GRAMELS, B.S. in E.E., New York University, 1936. Bell Telephone Laboratories 1925-. Mr. Gramels was first occupied with testing and development work in transmission, including handsets, recording apparatus and 33 rpm records. In 1937 and 1938 he worked on electrolytic condensers and silicon carbide varistors. Since 1938 he has been concerned with investigations of selenium rectifier cells both for Bell System applications and for military use. Member of the A.I.E.E.

MASON A. LOGAN, B.S. in Physics and Engineering, California Institute of Technology, 1927; M.A. in Physics, Columbia University, 1933. Carnegie Institute of California, Seismological Laboratory, 1926-1927. Bell Telephone Laboratories, 1927-. His early Laboratories' projects were concerned with wire transmission problems particularly those of losses, noise and cross induction in local, manual and dial telephone circuits. This was followed by circuit research on alternating current methods of signaling including the use of non-linear elements and electronic terminal equipment. From 1941 to 1948 he worked on military projects, including a mine fire control system, anti-aircraft gun director, magnetic proximity fuses, and guided missiles. For the past five years he has been a

member of the Switching Apparatus Development Department in which he is supervising a group concerned with static and dynamic behavior of new electromagnets and relays. He is also engaged in investigations of the performance of electrical contacts on telephone relays.

C. E. POLLARD, JR., Polytechnic Institute of Brooklyn, 1927-1931. Bell Telephone Laboratories, 1925-. Mr. Pollard first worked on voice recording and reproducing equipment and spent some time on the development of telephone carbon microphones before becoming interested in mercury contact relays. In this field he has been concerned with a wide variety of relays, and during World War II concentrated on their application to military projects. He is currently engaged in development work on mercury contact relays.

JOHN H. ROWEN, B.E.E., Ohio State University, 1948; M.S., Ohio State University, 1951. U.S.N.R., maintenance of Air Force radar equipment, 1944-1946; Ohio State University, Research Foundation, Antenna Laboratory, 1948-1951; Bell Telephone Laboratories, 1951-. Concerned with the practical application of fundamental research, he has spent his two years at Bell Laboratories on studies of the microwave behavior of ferrites. While at Ohio State University, Mr. Rowen worked on several methods of measuring the radiation efficiency of small aperture antennas. Member of the I.R.E. and Eta Kappa Nu.

MARK A. TOWNSEND, B.S. in E.E., Texas Technological College, 1936; S.M. in E.E., Massachusetts Institute of Technology, 1937. General Electric Company, 1937-1943; Massachusetts Institute of Technology, Radar School, 1943-1945. Bell Telephone Laboratories, 1945-. Mr. Townsend has been concerned with the basic development of gas-filled tubes. His projects have included work on the voltage reference tube, cold cathode stepping tubes, and the development of tubes for use in transmission and switching. At present Mr. Townsend is in charge of a group responsible for basic development of gas-filled tubes. Member of the A.I.E.E., Tau Beta Pi and Sigma Xi.

THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

ADVISORY BOARD

S. BRACKEN

F. R. KAPPEL

M. J. KELLY

EDITORIAL COMMITTEE

E. I. GREEN, *Chairman*

A. J. BUSCH

F. R. LACK

W. H. DOHERTY

W. H. NUNN

G. D. EDWARDS

H. I. ROMNES

J. B. FISK

H. V. SCHMIDT

R. K. HONAMAN

G. N. THAYER

EDITORIAL STAFF

J. D. TEBO, *Editor*

M. E. STRIEBY, *Managing Editor*

R. L. SHEPHERD, *Production Editor*

INDEX

VOLUME XXXII

1953

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

LIST OF ISSUES IN VOLUME XXXII

No. 1	January.....	Pages	1-264
" 2	March.....		265-522
" 3	May.....		523-778
" 4	July.....		779-1018
" 5	September.....		1019-1270
" 6	November.....		1271-1518

Index to Volume XXXII

A

- ADP *See* Ammonium dihydrogen phosphate
- AIEE *See* American Institute of Electrical Engineers
- AM *See* Amplitude modulation
- AMA *See* Automatic message accounting
- ABSTRACT(s) 255-60, 506-18, 767-74, 1007-14, 1257-65, 1507-13
- "Acceleration Effects on Electron Tubes" (F. W. Stubner) 1203-29
- "Acoustic Gyrator" (W. E. Kock)
abstract 1261
- Ahearn, A. J.
"Formative of Negative Ions of Sulfur Hexafluoride"
abstract 1007
- "Ähnlichkeit zwischen Vokalformanten und Formanten von Musikinstrumenten" (W. E. Kock) 1261
- "AIEE Progress" (D. A. Quarles)
abstract 1013
- Alley, Reuben, E. J., Jr.
biographical material 1267
"Review of New Magnetic Phenomena" 1155-72
- ALLOY(s)
copper, photometric determination of silicon 772
ferrous, silicon, photometric determination 772
magnetic properties 1155, 1258
Mishima, physical and magnetic structure 1262
nickel-carbon, reaction rates with barium oxide 771
silicon p-n junction diodes 512
zone-melting 513
- Alsberg, D. A.
"Principles and Applications of Converters for High-Frequency Measurements"
abstract 506
- AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
Quarles address, Phoenix meeting 514
telephone invention 768
- AMERICAN SOCIETY OF CIVIL ENGINEERS AND ARCHITECTS
centennial 768
professional engineering 514
- AMMONIUM DIHYDROGEN PHOSPHATE
dielectric properties 510
elastic properties 510
piezoelectric properties 510
polymorphism 257
properties interpretation 510
- AMPLIFIER(s)
L3 coaxial system 879-914
regenerative, for digital computers 508
transistor — cutoff frequency 516
traveling-wave-type, broad-band interdigital circuit for use in 256
vacuum tube electrometer 1261
- AMPLITUDE MODULATION
land mobile service 62
- "Analysis of Measurements on Magnetic Ferrites" (C. D. Owens)
abstract 1012
- Anderson, A. E.
"Transistors in Switching Circuits"
abstract 506-7
- Anderson, F. B.
"Gain and Phase Angle Measuring Set"
abstract 1007
- Anderson, J. R.
"Electrical Delay Lines for Digital Computer Applications"
abstract 1507
"Ferroelectric Storage Elements for

Anderson, J. R.—*continued*:

Digital Computers and Switching Systems"

abstract 506

Anderson, P. W.

"Concept of Spin-Lattice Relaxation in Ferromagnetic Materials" 767

"Exchange Narrowing in Paramagnetic Resonance"

abstract 1257-8

Andrus, J.

"Motion of Domain Walls in Ferrite Crystals" 1260

ANGLO-AMERICAN COMMITTEE ON TECHNICAL TERMINOLOGY

report 773

"Anisotropic Elasticity with Applications to Dislocation Theory" (Eshelby, Read and Shockley)

abstract 1509

ANSWERING SERVICES *See* Telephone answering services

ANTENNA(S)

ground-plane, matching coax line 258

microwave relay stations 770

-reflector problem, theoretical study 770

ANTI-FERROMAGNETIC COBALT DISCLINITY

domain structure evidence 508

ANTIMONY

photometric determination in lead 1261

"Application of Information Theory to Research in Experimental Phonetics" (G. E. Peterson)

abstract 513

"Approximating the Mode From Weighted Sample Values" (H. L. Jones)

abstract 1010

ARC, short, characteristics 1493-1506

"Arcing of Electrical Contacts in Telephone Switching Circuits" (M. M. Atalla)

Part I — Theory of the Initiation of the Short Arc 1231-44

Part II — Characteristics of the Short Arc 1493-1506

"Arithmetic Processes for Digital Computers" (J. H. Felker)

abstract 1009

Armstrong, C. A.

"Communications for Civil Defense" 1507

abstract 1007

"Telephone Industry in National Defense"

abstract 507

Arnold, S. M.

"Growth of Metal Whiskers" 1261

"ASTM Standards — Their Effect on Plastics Technology" (R. Burns)

abstract 255

Atalla, M. M.

"Arcing of Electrical Contacts in Telephone Switching Circuits"

Part I — "Theory of the Initiation of the Short Arc" 1231-44

Part II — "Characteristics of the Short Arc" 1493-1506

biographical material 1267

ATOM(S)

ionization, single, by electrons, threshold law 1512-13

magnetic resonance 74, 384

"Attenuation Equalizers" (F. R. Bies)

abstract 1258

AUDIO FREQUENCY SIGNALS

transmission, cold cathode tubes for 1371-91

"Auditory Tests with Synthetic Vowels" (R. L. Miller)

abstract 1011

"Automatic Line Insulation Test Equipment for Local Crossbar Systems" (Burns and Dehn) 627-46

AUTOMATIC MESSAGE ACCOUNTING

tape-to-card conversion 1260

toll messages billing, 1264

"Automatic Recognition of Spoken Digits" (Davis, Biddulph and Bala-shuk)

abstract 768-9

"Automatic Switching for Nation-Wide Telephone Service" (Clark and Osborne)

abstract 507

"Automatic Toll Switching Systems"
(F. F. Shipley) 514

B

Babcock, Wallace C.

biographical material 261

"Intermodulation Interference in
Radio Systems" 63-73

BAKELITE (photoelastic)

connections, solderless, wrapped 557

"Balanced Polar Mercury Contact Re-
lay" (Brown and Pollard) 1393-
1411

Balashek, S.

"Automatic Recognition of Spoken
Digits"

abstract 768-9

BAND(s)

radio, mobile, frequency economy 42-
62

Bardeen, John

biographical material 261

"Surface Properties of Germanium"
1-41

BARIUM OXIDE

nickel-carbon alloys, reaction rates
771

BARIUM TITANATE

domain properties 511

Barotta, P. J.

"Transistor Physics Simplified" 1510

"Basic Physical Systems for Communi-
cation Between Two Individuals"
(G. E. Peterson) 1511

"Basic Theorems of Information
Theory" (B. McMillan)

abstract 1510

BATTERY(ies)

lead-acid, stationary 257

Beach, A. L.

"Solubility and Diffusion Coefficient
of Carbon in Nickel: Reaction
Rates of Nickel-Carbon Alloys
with Barium Oxide"

abstract 771

Beck, A. C.

"Low-Loss Waveguide Transmission"
abstract 1011

Becker, J. A.

"Field Emission Microscope and Flash
Filament Techniques for the Study
of Structure and Adsorption on
Metal Surfaces"

abstract 1008

"Behavior of Magnetic Materials" (R.
M. Bozorth)

abstract 1259

BELL TELEPHONE LABORATORIES TRAN-
SISTOR TEACHERS SUMMER SCHOOL
767

BEND(s)

circular electric wave, transmission
around 511

Benedict, T. S.

"Microwave Observation of the Col-
lision Frequency of Electrons in
Germanium" 1258

Bennett, A. F.

biographical material 775

"Improved Circuit for the Telephone
Set" 611-26

Bennett, William R.

biographical material 1267-8

"Correlatograph — A Machine for
Continuous Display of Short Term
Correlation" 1173-85

Biddulph, R.

"Automatic Recognition of Spoken
Digits"

abstract 768-9

Bies, F. R.

"Attenuation Equalizers"

abstract 1258

BIOASSAY TEST 152, 154

Birdsall, H. A.

"Microstructure of Capacitor Paper"
1510

Black, H. S.

"Experimental Verification of the
Theory of Laminated Conductors"
abstract 255

Blount, Frank E.

biographical material 1514

"Transistor Oscillator for Use in
Multifrequency Pulsing Current
Supply" 1313-31

- Bodle, D. W.
 "Lightning Protection for Mobile Radio Fixed Stations"
 abstract 257
- Bogert, B. P.
 "Band Width of Vowel Formants"
 abstract 1507-8
- BOLTZMANN EQUATION 170, 193
 "Bombardment Conductivity" (K. G. McKay) 1510
- BOOLEAN FUNCTIONS
 symmetry types, on n variables 1264
- Bozorth, R. M.
 "Behavior of Magnetic Materials"
 abstract 1259
 "Magnetic Crystal Anisotropy and Magnetostriction of Iron-Nickel Alloys"
 abstract 1008-9
 "Magnetic Study of Low Temperature Transformation in Magnetite" 1264
 "Measurement of Magnetostriction in Single Crystals"
 abstract 1008
 "Permalloy Problem"
 abstract 1258-9
- Brangaccio, D. J.
 "Factors Affecting Traveling Wave Tube Power Capacity" 1509
- Brattain, Walter H.
 biographical material 261
 "Electron Spin Resonance in a Silicon Semi-Conductor" 1511
 "Surface Properties of Germanium" 1-41
- Bridgman, D. C.
 "College Graduates and the Country's Telephone Industry" 767
- Briggs, H. B.
 "Infrared Absorption in High Purity Germanium"
 abstract 507
 "New Infrared Absorption Bands in p-Type Germanium"
 abstract 507
 "Broad-Band Interdigital Circuit for Use in Traveling-Wave-Type Amplifiers" (R. C. Fletcher)
 abstract 256
- "Broad-Band Matching with a Directional Coupler" (W. C. Jakes)
 abstract 509-10
- Brown, J. T. Lindsay
 biographical material 1514
 "Balanced Polar Mercury Contact Relay" 1393-1411
- Buehler, E.
 "Single-Crystal Germanium"
 abstract 257
- Bullington, Kenneth
 biographical material 262
 "Frequency Economy in Mobile Radio Bands" 42-62
 "Frequency Economy in Mobile Radio Bands," IRE Trans. 1508
 "Radio Transmission Beyond the Horizon in the 40- to 4,000-Mc Band"
 abstract 767-8
- Burns, R.
 "ASTM Standards — Their Effect on Plastics Technology"
 abstract 255
- Burns, R. M.
 "Science and Scientists in Telecommunications" 1259
- Burns, R. W.
 "Automatic Line Insulation Test Equipment for Local Crossbar Systems" 627-46
 biographical material 775
- C
- CABLE(s)
 Clogston 695 ff
 coaxial:
 Clogston Cable compared with 703-5
 See also carrier: L3
 multi-conductor 1245
 polyethylene insulation 1245
- Calbick, C. J.
 "Microstructure of Capacitor Paper" 1510
 "Calibration of the Rolling Ball Viscometer" (H. W. Lewis) 1010
- CALL(s)

- delay curves for calls served at random, Riordan 100-19, 1266
 delayed exponential, served in random order, working curves for, Wilkinson 360-83
 queueing 367
 CAMERA(s)
 three-phase power from single-phase source 256
 Campbell, R. D.
 "Path Testing for Microwave Radio Routes" 1508
 Campbell, W. E.
 "Solid Lubricants" 1508
 CAPACITOR DIELECTRICS 1264-5
 CAPACITOR PAPER
 microstructure 1510
 CARBON
 (in) nickel, solubility and diffusion coefficient 771
 CARBON MONOXIDE
 ions, drift velocity 1013
 CARD TRANSLATOR *See* Translator
 "Card Translator for Nationwide Dialing" (Hampton and Newsom) 1037-98
 CARRIER(s), CARRIER SYSTEMS
 L3:
 amplifiers 879-914
 equalization 833-878, 943
 manufacture 969
 quality control 943-967, 969
 regulation 833-878
 system design 779-832
 television terminals 915-942
 Type-N 1259
 Type-O 1259
 CAUSE (assignable)
 type detection 512
 CAVITATION
 water, ultrasonically induced 1513
 "Charge Transfer and the Mobility of Rare Gas Ions" (J. A. Hornbeck)
 abstract 509
 Ciccolella, D. F.
 "High Frequency Crystal Units for Use in Selective Networks and Their Proposed Application in Filters Suitable for Mobile Radio Channel Selection" 1508
 CIRCUIT(s)
 telephone set (500 type) 611-26
 CIVIL DEFENSE
 communication 1007, 1507
 telephone system 258
 Clark, A. B.
 "Automatic Switching for Nationwide Telephone Service"
 abstract 507
 "Telephony Development in the United States"
 abstract 768
 Clarke, K. B.
 "Western Electric's Service with Standards"
 abstract 507
 CLOGSTON TYPE CONDUCTORS *See* Conductor
 Clos, Charles
 biographical material 519
 "Non-Blocking Switching Networks Study" 406-24
 COAXIAL CARRIER SYSTEM L3 *See* Carrier
 COBALT DISILICIDE
 superconducting properties 1011
 COBALT OXIDE (CoO)
 domain structure evidence 508
 magnetomechanical effects 1260
 COBALT-SILICON SYSTEM
 superconductivity 256
 COLD
 pressure connection 533
 COLD-CATHODE STEPPING TUBE *See* Electron tube
 "Cold Cathode Tubes for Transmission of Audio Frequency Signals" (Townsend and Depp) 1371-91
 "College Graduates and the Country's Telephone Industry" (D. C. Bridgman) 767
 Colley, Reginald H.
 biographical material 262
 "Evaluation of Wood Preservatives" 120-69, 425-505
 "Review of American Standard Fiber Stresses of Wood Poles" 1262

COMMUNICATION

- between two individuals, physical systems 1511
- "Communications for Civil Defense" (C. A. Armstrong) 1507
- abstract 1007
- "Comparison of Mobile Radio Transmission at 150, 450, 900 and 3700 Mc" (W. R. Young) 1513
- "Comparison of Recording Processes" (J. G. Frayne)
- abstract 509

COMPUTER(S)

- binary, functions 769
- digital:
 - delay lines for 1507
 - ferroelectric storage elements 506
 - regenerative amplifier 508
 - transistor, 255, 508, 769
- historical development 772
- "Computers — Past, Present, and Future" (W. H. MacWilliams, Jr.)
- abstract 772

CONDUCTIVITY

- bombardment 1510
- electric fields, high, mobility in 768

CONDUCTOR(S)

- Clogston type, laminated, transmission properties 695-713
- electroformed for drop wire 1099-1135
- laminated 695
- experimental verification of the theory of 255

CONNECTION(S)

- permanent, conditions for obtaining 551-91
- pressure 525, 526
- quality 532
- solder 525
- solderless, wrapped 523-610
- "Connection Formulas Between the Solutions of Mathieu's Equation" (G. H. Wannier)
- abstract 1014

CONTACT(S) *See* Electrical contacts

CONTACT EROSION

- switching circuits 1493
- "Controlled Gas Leak" (J. Morrison) 1262

CONVERTER(S)

- high-frequency measurements 506
- transistor, negative-impedance 1510
- "Conveyorization Avoids Handling Headaches" (P. T. Mathy) 1262

Conwell, Esther M.

- "High Field Mobility in Germanium With Impurity Scattering Dominant"
- abstract 1508
- "Mobility in High Electric Fields"
- abstract 768
- "Mobility of Electrons in Germanium"
- abstract 507
- "Properties of Silicon and Germanium"
- abstract 507

CoO *See* Cobalt oxide

COPPER ALLOYS

- silicon, photometric determination 772
- "Copper as an Acceptor Element in Germanium" (Fuller and Struthers)
- abstract 256

COPPER OXIDE

- varistors 514

COPPER-STEEL WIRE *See* Wire

- "Correlatograph — A Machine for Continuous Display of Short Term Correlation" (W. R. Bennett) 1173-85

CORROSION

- connections, solderless, wrapped 598

COUNTER(S)

- crystal conduction 1262
- reversible binary 516
- "Coupled Resonator Reflex Klystron" (E. D. Reed) 715-66

COUPLER (directional)

- broad band matching 509
- multi-element 258, 511

Coy, J. A.

- "Type-O Carrier Telephone"
- abstract 1259

CREOSOTE

- as wood preservative 458, 461, 469, 473, 474, 479, 489-90

CROSSBAR SYSTEM No. 1

- automatic line insulation testing 627-46

- CROSSBAR SYSTEM No. 5
 automatic line insulation testing 627-46
 throwdown machine 292-359
 traffic-carrying characteristics study 292-359
- CROSSTALK
 interchannel 514
- CRYSTAL(s)
 antiferroelectric 510
 conduction counter 1262
 dislocated, geometrical relations 1012
 electrons, e/m values, interpretation 515
 ferrites, motion of domain walls 1260
 germanium: single, purification and segregations prevention 1012
 solute distribution 1512
 high-frequency units for primary frequency standards 260
 high-frequency units for use in selective networks and proposed application in filters for mobile radio channel selection 1508
 magnetic resonance 74
 "Crystal Conduction Counter" (K. G. McKay)
 abstract 1262
- Cutler, C. C.
 "Factors Affecting Traveling Wave Tube Power Capacity" 1509
- D
- Dacey, G. C.
 "Space-Charge Limited Hole Current in Germanium"
 abstract 1509
- Dalton, A. G.
 "Practice of Quality Control"
 abstract 1009
- Darrow, Karl K.
 biographical material 262, 519-20
 "Magnetic Resonance of Electrons" 384-405
 "Nuclear Magnetic Resonance" 74-99
- Davis, K. H.
 "Automatic Recognition of Spoken Digits"
 abstract 768-9
- "DC Field Distribution in a 'Swept Intrinsic' Semiconductor Configuration" (R. C. Prim) 665-94
- Debye, P. P.
 "Mobility of Electrons in Germanium"
 abstract 507
- DeCamp, R. T.
 "Matching Coax Line to the Ground-Plane Antenna"
 abstract 258
- DECIBEL 6
 in search of the missing 512
- DEFENSE *See* Civil defense
- Dehn, J. W.
 "Automatic Line Insulation Test Equipment for Local Crossbar Systems" 627-46
 biographical material 775
- DELAY CALLS *See* Call
- "Delay Curves for Calls Served at Random" (John Riordan) 100-19
 addendum to 1266
- DELAY FORMULA *See* Erlang
- DELAY LINES
 digital computers 1507
 storage capacity 1507
- Depp, Wallace A.
 biographical material 1514-15
 "Cold Cathode Tubes for Transmission of Audio Frequency Signals" 1371-91
- "Design Theory of Junction Transistors" (J. M. Early) 1271-1312
- "Development and Manufacture of Electroformed Conductor for Telephone Drop Wire" (Gray and Murray) 1099-1135
- DIAL TELEPHONE, DIALING
 card translator for nationwide 1037
 exchanges, traffic engineering design 257, 516
 numbering system, nation-wide, need for 512
- Dickinson, Frank R.
 biographical material 1015
 "L3 Coaxial System — Amplifiers" 879-914

- Dickten, E.
 "High-Frequency Transistor Tetrode"
 abstract 774
 "Junction Transistor Tetrode for
 High-Frequency Use"
 abstract 517
- DIELECTRIC(s)
 capacitor 1264-5
 $\text{ND}_4\text{D}_2\text{PO}_4$, properties 510
 "Diffusion of Lithium Into Germanium
 and Silicon" (Fuller and Ditz-
 enberger) 1509
- DIGIT(s)
 register 508-9
 spoken, automatic recognition 768-9
- DIGITAL COMPUTERS *See* Computer
- DIMENSIONAL RESONANCE *See* Reso-
 nance
- DIODE(s)
 semiconductor gates 1137-54
 silicon p-n junction alloy 512
- DIRECTIONAL COUPLER *See* Coupler
- DISLOCATION THEORY
 anisotropic elasticity 1509
- Ditzenberger, J. A.
 "Diffusion of Lithium Into Germanium
 and Silicon" 1509
- Dodge, H. F.
 biographical material 1015
 "L3 Coaxial System — Quality Con-
 trol Requirements" 943-67
 "Domain Properties in BaTiO_3 " (W. J.
 Merz)
 abstract 511
- DOMAIN STRUCTURE
 anti-ferromagnetic CoO 508
- DOMAIN WALLS
 ferrite crystals 1260
 motion 1162, 1260
 "Dominant Wave Transmission Char-
 acteristics of a Multimode Round
 Waveguide" (A. P. King)
 abstract 256
 "Drift Velocities of Ions in Krypton and
 Xenon" (R. N. Varney)
 abstract 517
 "Drift Velocity of Ions in Oxygen, Nitro-
 gen, and Carbon Monoxide" (R. N.
 Varney)
 abstract 1013
- DROP WIRE *See* Wire
- Drvostep, J. J.
 "Standardization of Rigid Coaxial
 Transmission Lines" 1009
 "Dual Photomagnetic Intermediate
 Studio Recording" (Frayne and
 Livadary)
 abstract 769
 "Dynamic Measurements on Electro-
 magnetic Devices" (M. A. Logan)
 1413-67
 "Dynamic Program for Conversion"
 (J. R. Townsend) 1512
 "Dynamic Spectrograms of Speech"
 (Kock and Miller) 771
 "Dynamics of Transistor Negative-Re-
 sistance Circuits" (B. G. Farley)
 abstract 508
- E**
- EAR CANAL
 sound pressure 512
- Early, James M.
 biographical material 1515
 "Design Theory of Junction Trans-
 istors" 1271-1312
 "Effects of Space-Charge Layer
 Widening in Junction Transistors"
 abstract 507-8
- Ebers, J. J.
 "Four-Terminal p-n-p-n Transistors"
 abstract 508
- ECASS *See* Electronically controlled
 automatic switching system
- ECHO(es)
 television transmission 1511
 "Economics of High-Speed Photogra-
 phy" (A. C. Keller)
 abstract 771
 "Effect of Electrode Spacing on the
 Equivalent Base Resistance of
 Point-Contact Transistors" (L. B.
 Valdes)
 abstract 516
 "Effects of Space-Charge Layer Widen-

- ing in Junction Transistors" (J. M. Early)
 abstract 507-8
- Ehrbar, R. D.
 biographical material 1015
 "L3 Coaxial System — System Design" 781-832
- ELASTICITY
 anisotropic, applications to dislocation theory 1509
 antiferromagnetic CoO, evidence of domain structure, measurements for 508
 $\text{ND}_4\text{D}_2\text{PO}_4$, properties 510
 "Elasticity and Thermal Expansion of Germanium between -195 and 275°C " (M. E. Fine)
 abstract 1260
- ELECTRIC FIELDS (high)
 mobility 768
- ELECTRIC WAVE
 circular, transmitting around bends 511
- ELECTRICAL CONTACTS
 switching circuits, arcing of 1493-1506
 "Electrical Delay Lines for Digital Computer Applications" (J. R. Anderson)
 abstract 1507
- ELECTRICAL STABILITY
 connections, solderless, wrapped 591
- ELECTRODE SPACING
 equivalent base resistance of 516
 transistors, point-contact 516
 "Electroformed Copper-steel Wire Development" (A. N. Gray)
 abstract 1260
- ELECTROLYSIS SWITCH
 improved 259
- ELECTROMAGNETIC DEVICES
 dynamic measurements 1413-67
- ELECTROMETER AMPLIFIER
 vacuum tube 1261
- ELECTRON(s)
 crystals, e/m values, interpretation 515
 ejection from Mo by He^+ , He^{++} , and He_2^+ 769
 electric fields, high, mobility in 768
 germanium:
 collision frequency, microwave observation 1258
 mobility in 507
 magnetic resonance 74, 384-405
 mean free paths in evaporated metal films 514
 microscopy, of solids 770
 mobility in high electric fields 1511-12
 recombination statistics 259
 resonance, *See* Electron — magnetic resonance
 spin resonance in a silicon semiconductor 1511
 streams, microwave noise calculation 773
 threshold law for single ionization of atoms or ions 1512-13
- ELECTRON(S) AND HOLES
 diffusion constant and mobility 767
 mobility in high electric fields 1511-12
 recombination statistics 259
 "Electron Ejection from Mo by He^+ , He^{++} , and He_2^+ " (H. D. Hagstrum)
 abstract 769-70
 "Electron Spin Resonance in a Silicon Semi-Conductor" (Portis, Kip, Kittel and Brattain) 1511
- ELECTRON TUBE(s)
 acceleration effects 1203-29
 cold cathode, for transmission of audio frequency signals 1371-91
 computer designs 769
 electrometer amplifier 1261
 shock measurement 1203, 1215, 1219-21
 stepping, cold-cathode, 10-stage 773
 traveling wave 1510
 factors affecting power capacity 1509
- ELECTRONICALLY CONTROLLED AUTOMATIC SWITCHING SYSTEM 406
- ELECTRONICS
 solid state physics 257
- Elmendorf, C. H.
 biographical material 1016
 "L3 Coaxial System — System Design" 781-832

ENCODER

optical position 508

ENGINEERING PROFESSION

organization 514

EQUALIZER(s)

attenuation 1258

EQUATION(s)

transistor 1013

Erlang, A. K. 100

ERLANG FORMULA 100, 1266

EROSION *See* Contact Erosion

Eshelby, J. D.

"Anisotropic Elasticity with Applications to Dislocation Theory"

abstract 1509

"Evaluation of Wood Preservatives" (Reginald H. Colley) 120-69, 425-505

"Evidence for Domain Structure in Antiferromagnetic CoO From Elasticity Measurements" (M. E. Fine)

abstract 508

"Exchange Narrowing in Paramagnetic Resonance" (Anderson and Weiss)

abstract 1257-8

"Experimental Verification of the Relationship Between Diffusion Constant and Mobility of Electrons and Holes" 767

abstract

"Experimental Verification of the Theory of Laminated Conductors" (Black, Mallinckrodt and Morgan)

abstract 255

F

FM *See* Frequency modulation

"Factors Affecting Traveling Wave Tube Power Capacity" (Cutler and Braggaccio) 1509

FADING

radio signals, super-high frequency 1187-1202

FARADAY EFFECT

ferromagnetic 1260, 1333

microwave 1166

Farley, B. G.

"Dynamics of Transistor Negative-Resistance Circuits"

abstract 508

Felker, J. H.

"Arithmetic Processes for Digital Computers"

abstract 1009

"Regenerative Amplifier for Digital Computer Applications"

abstract 508

"Typical Block Diagrams for a Transistor Digital Computer"

abstract 255, 769

FEMF (foreign e.m.f.) TEST 636

FERRITE(s)

crystals, motion of domain walls 1260

description 1156

gyrator 1261

inductor cores 265-91

initial permeability and related losses 256

magnetic, measurements analysis 1012

magnetic phenomena 1155 ff

waveguide medium, new non-reciprocal 1512

"Ferrite Core Inductors" (H. A. Stone, Jr.) 265-91

"Ferrites in Microwave Applications" (J. H. Rowen) 1333-69

"Ferroelectric Storage Elements for Digital Computers and Switching Systems" (J. R. Anderson)

abstract 506

FERROMAGNET

spherical model 510

FERROMAGNETIC ALLOYS

silicon, photometric determination 772

"Ferromagnetic Faraday Effect at Microwave Frequencies and Its Applications" (C. L. Hogan) 1260

FERROMAGNETIC RESONANCE 1164-6

FERROMAGNETIC SUBSTANCES

magnetic resonance 74, 398

relaxation, spin-lattice 767

FERROUS ALLOYS

silicon, photometric determination 772

FIELDS *See* Electric Fields; Magnetic Field

"Field Emission Microscope and Flash

- Filament Techniques for the Study of Structure and Adsorption on Metal Surfaces" (Becker and Hartman)
abstract 1008
- FILAMENTARY TRANSISTORS *See* Transistor
- FILM(s), metal, evaporated
electron mean free paths 514
- FILM PULLING MECHANISMS
three-phase power from single-phase source 256
- FILTER(s)
mobile radio channel selection, high frequency crystal units 1508
- Finch, Tudor R.
biographical material 1016
"L3 Coaxial System — Equalization and Regulation" 833-78
- Fine, M. E.
"Elasticity and Thermal Expansion of Germanium between -195 and 275°C"
abstract 1260
"Evidence for Domain Structure in Antiferromagnetic CoO From Elasticity Measurements"
abstract 508
"Magnetomechanical Effects in an Antiferromagnet, CoO" 1260
- Fletcher, R. C.
"Broad-Band Interdigital Circuit for Use in Traveling-Wave-Type Amplifiers"
abstract 256
"New Infrared Absorption Bands in *p*-Type Germanium"
abstract 507
- Follingstad, H. G.
"Optical Position Encoder and Digit Register"
abstract 508-9
"Formative of Negative Ions of Sulfur Hexafluoride" (Ahearn and Hannay)
abstract 1007
"Four-Terminal *p*-*n*-*p*-*n* Transistors" (J. J. Ebers)
abstract 508
- Fox, A. G.
"Magnetic Double Refraction at Microwave Frequencies" 517
- Frayne, J. G.
"Comparison of Recording Processes"
abstract 509
"Dual Photomagnetic Intermediate Studio Recording"
abstract 769
"Frequency Economy in Mobile Radio Bands" (Kenneth Bullington) 42-62, 1508
- FREQUENCY MODULATION
land mobile service 62
"Frequency Response and Stability of Point-Contact Transistors" (B. N. Slade)
abstract 515
- Frost, George R.
biographical material 520
"Throwdown Machine for Telephone Traffic Studies" 292-359
- Fuller, C. S.
"Copper as an Acceptor Element in Germanium"
abstract 256
"Diffusion of Lithium Into Germanium and Silicon" 1509
"Properties of Thermally Produced Acceptors in Germanium"
abstract 256
- G**
- "Gain and Phase Angle Measuring Set" (Anderson)
abstract 1007
- Galt, J. K.
"Initial Permeability and Related Losses in Ferrites"
abstract 256
"Motion of Domain Walls in Ferrite Crystals" 1260
- Garrett, Robert F.
biographical material 1016
"L3 Coaxial System — Application of Quality Control Requirements in the Manufacture of Components" 969-1005
- GAS(es), rare

- Gas(es)—*continued*:
 ion charge transfer and mobility 509
 ion-atom collisions 509
- GAS LEAK
 controlled 1262
- GASEOUS IONS
 motion in strong electric fields 170-254, 259
- GASEOUS KINETICS
 charged particles moving through a gas under the influence of a static, uniform electric field 170
 equations 178
 glossary 178
- Gentile, S. P.
 "Transistor Physics Simplified" 1510
 "Geometrical Relations in Dislocated Crystals" (J. F. Nye)
 abstract 1012
- GERMANIUM
 acceptors, thermally produced, properties 256
 carrier density and mobility 507
 copper as an acceptor element 256
 crystals:
 single 257
 solute distribution 1512
 uses for the Hall effect 1010
 elasticity 1260
 electrical properties 507
 electron mobility 507
 electrons, collision frequency, microwave observation 1258
 electrons and holes:
 diffusion constant and mobility, relationship 767
 field dependence of mobility 1511
 filament, theory of magnetic effects on the noise in a 647-64
 high field mobility in, with impurity scattering dominant 1508
 high purity, infrared absorption 507
 infrared absorption 507
 lithium diffusion 1509
 minority carrier lifetime, measurement 517
 phototransistors 1263
 properties 507
 p-type, infrared absorption bands 507
 space-charge limited hole current 1509
 surface properties 1-41
 temporary traps 1260
 thermal conversion, impurity effects 257
 thermal expansion 1260
 varistors 514
- Goucher, F. S.
 "Interpretation of α -values in *p-n* Junction Transistors"
 abstract 1009
- Graham, R. Shiels
 biographical material 1016
 "L3 Coaxial System — Television Terminals" 915-42
- Gramels, Joseph
 biographical material 1515
 "Selenium Rectifiers — Factors in Their Application" 1469-92
- Gray, A. N.
 biographical material 1268
 "Development and Manufacture of Electroformed Conductor for Telephone Drop Wire" 1099-1135
 "Development of Electroformed Copper-Steel Wire"
 abstract 1260
- Green, E. I.
 biographical material 1016-17
 "L3 Coaxial System — Foreword" 779-80
- GREENSALT
 evaluation 487
- Grossman, Alexander J.
 biographical material 1017
 "L3 Coaxial System — System Design" 781-832
- Groth, W. B.
 "Principles of Tape-to-Card Conversion in the AMA System" 1260
 "Growth of Metal Whiskers" (Koonce and Arnold) 1261
- GYRATOR(s)
 acoustic waves 1261
 employing magnetic field independent orientations in germanium 1010
 ferrite 1261

H

- Hagstrum, H. D.
 "Electron Ejection from Mo by He^+
 He^{++} , and He_2^{++} "
 abstract 769-70
- "Hall Effect Modulators and 'Gyrators'
 Employing Magnetic Field Inde-
 pendent Orientations in German-
 ium" (Mason, Hewitt, and Wick)
 abstract 1010
- Hamming, R. W.
 "Measurement of Magnetostriction in
 Single Crystals"
 abstract 1008
- Hampton, L. N.
 biographical material 1268
- "Card Translator for Nationwide
 Dialing" 1037-98
- Hannay, H. B.
 "Formative of Negative Ions of Sulfur
 Hexafluoride"
 abstract 1007
- "Hard Rubber" (H. Peters)
 abstract 513
- Harris, J. R.
 "Transistor Shift Register and Serial
 Adder"
 abstract 509
- Hartman, C. D.
 "Field Emission Microscope and Flash
 Filament Techniques for the
 Study of Structure and Adsorp-
 tion on Metal Surfaces"
 abstract 1008
- Haynes, J. R.
 "Temporary Traps in Silicon and Ger-
 manium" 1260
- HEAT
 pressure connection 533
 wood sterilization 456
- Heidenreich, R. D.
 "Methods in Electron Microscopy of
 solids"
 abstract 770
- "Physical and Magnetic Structure of
 the Mishima Alloys" 1262
- HELIUM ISOTOPES
 vapor pressures 772

Hewitt, W. H.

"Hall Effect Modulators and 'Gyra-
 tors' Employing Magnetic Field
 Independent Orientations in Ger-
 manium"

abstract 1010

"High Field Mobility in Germanium
 With Impurity Scattering Dom-
 inant" (E. M. Conwell)

abstract 1508

HIGH-FREQUENCY

measurements, converters for 506

"High-Frequency Crystal Units for
 Primary Frequency Standards" (A.
 W. Warner)

abstract 260

"High-Frequency Crystal Units for Use
 in Selective Networks and Their
 Proposed Application in Filters Suit-
 able for Mobile Radio Channel Se-
 lection" (Ciccolella and Labrie)
 1508

"High-Frequency Transistor Tetrode"
 (Wallace, Schimpf and Dickten)

abstract 774

HIGH-SPEED PHOTOGRAPHY *See* Photog-
raphy

"Higher Frequencies for Ground-Air
 Communications" (S. B. Wright)
 1265

Hines, M. E.

"Traveling-Wave Tube" 1510

Hogan, C. L.

"Ferromagnetic Faraday Effect at
 Microwave Frequencies and Its Ap-
 plications" 1260

Holcomb, A. L.

"Three-Phase Power From Single-
 Phase Source"

abstract 256

HOLE(S) AND ELECTRONS *See* Electron(s)
 and holes

Hopper, A. L.

"Nonsynchronous Time Division with
 Holding and with Random Sam-
 pling"

abstract 258, 513

- Hopper, H. G.
 "Motion of Domain Walls in Ferrite Crystals" 1260
- Hornbeck, J. A.
 "Charge Transfer and the Mobility of Rare Gas Ions"
 abstract 509
- "Temporary Traps in Silicon and Germanium" 1260
- "How to Conceal Telephone Wires, Keep Desks Neat" (A. R. Hutchinson) 770
- "How To Detect the Type of an Assignable Cause" (P. S. Olmstead)
 abstract 512
- Hughes, W. T.
 "Vacuum Tube Electrometer Amplifier" 1261
- Hulm, J. K.
 "Superconducting Properties of Cobalt Disilicide"
 abstract 1011
- HUMIDITY**
 connections, solderless, wrapped 598
- Huntley, H. R.
 biographical material 1268
 "Transmission Design of Intertoll Telephone Trunks" 1019-36
- Hussey, Luther W.
 biographical material 1269
 "Semiconductor Diode Gates" 1137-54
- Hutchinson, A. R.
 "How to Conceal Telephone Wires, Keep Desks Neat" 770
- HYPERFINE STRUCTURE OF ELECTRON RESONANCE** 396, 405
- I**
- "Improved Circuit for the Telephone Set" (A. F. Bennett) 611-26
- "Improved Electrolysis Switch" (V. B. Pike)
 abstract 259
- "Impurity Effects in the Thermal Conversion of Germanium" (Slichter and Kolb)
 abstract 257
- "In Search of the Missing 6 Db" Munson and Wiener)
 abstract 512
- INDIUM ANTIMONY (InSb)**, magneto-resistance effect in 1263
- INDUCTOR(s)**
 ferrite materials as cores 265-91
- "Influence of Echoes on Television Transmission" (P. Mertz) 1511
- INFORMATION THEORY**
 basic theorems 1510
- "Information-Bearing Elements of Speech" (G. E. Peterson)
 abstract 773
- "Infrared Absorption in High Purity Germanium" (H. B. Briggs)
 abstract 507
- "Initial Permeability and Related Losses in Ferrites" (J. K. Galt)
 abstract 256
- InSb** See Indium antimony
- INSULATION**
 crossbar systems, local, automatic line test equipment 627-46
 leakage paths 627, 628
 moisture 627
 polyethylene 1245-56
 stress relaxation 517-18
 subscribers' lines 627
- "Interesting Property of Certain Conductive Rubbers" (L. G. Kersta) 1261
- INTERFERENCE**
 co-channel 42, 45
 intermodulation 63-73
- "Nonsynchronous Time Division with Holding and with Random Sampling" (Pierce and Hopper)
 abstract 258, 513
- "Intermodulation Interference in Radio Systems" (Wallace C. Babcock) 63-73
- "Interpretation of α -values in p-n Junction Transistors" (Goucher and Prince)
 abstract 1009
- "Interpretation of e/m Values for Electrons in Crystals" (W. Shockley) 515

ION(s)

- carbon monoxide, drift velocity 1013
- gaseous, motion in strong electric fields 259, 170-254
- gases, rare, charge transfer and mobility 509
- ionization, single, by electrons, threshold law 1512-13
- krypton, drift velocity 517
- nitrogen, drift velocity 1013
- oxygen, drift velocity 1013
- sulfur hexafluoride, negative 1007
- xenon, drift velocities 517

IRON-NICKEL ALLOYS

- magnetic properties 1258

J

Jakes, W. C., Jr.

- "Broad Band Matching with a Directional Coupler" abstract 509-10
- "Theoretical Study of an Antenna-Reflector Problem" abstract 770

JOINING METHODS *See* Connection

Jones, H. L.

- "Approximating the Mode From Weighted Sample Values" abstract 1010

JUNCTION TRANSISTOR(S) *See* Transistor

"Junction Transistor" (M. Sparks) abstract 516

"Junction Transistor Tetrode for High-Frequency Use" (Wallace, Schimpf and Dickten) abstract 517

K

Kaplan, E. L.

- "Tensor Notation and the Sampling Cumulants of k -Statistics" abstract 770

Kaylor, Robert L.

- biographical material 1269
- "Statistical Study of Selective Fading of Super-High Frequency Radio Signals" 1187-1202

Keister, William

- biographical material 520
- "Throwdown Machine for Telephone Traffic Studies" 292-359

Keller, A. C.

- "Economics of High-Speed Photography" abstract 771
- "New General-Purpose Relay for Telephone Switching Systems" abstract 510

Kern, H. E.

- "Solubility and Diffusion Coefficient of Carbon in Nickel: Reaction Rates of Nickel-Carbon Alloys with Barium Oxide" abstract 771

Kersta, L. G.

- "Interesting Property of Certain Conductive Rubbers" 1261

Ketchledge, R. W.

- biographical material 1017
- "L3 Coaxial System — Equalization and Regulation" 833-78

KINETIC ENERGY

- electrons 769

KINETICS *See* Gaseous kinetics

King, A. P.

- "Dominant Wave Transmission Characteristics of a Multimode Round Waveguide" abstract 256

Kinsburg, Boris J.

- biographical material 1017
- "L3 Coaxial System — Quality Control Requirements" 943-67

Kip, A. F.

- "Electron Spin Resonance in a Silicon Semi-Conductor" 1511

Kittel, C.

- "Electron Spin Resonance in a Silicon Semi-Conductor" 1511

Klie, Robert H.

- biographical material 1017
- "L3 Coaxial System — System Design" 781-832

KLYSTRON

- reflex, coupled resonator (tube) 715-66

- Kock, W. E.
 "Acoustic Gyrator"
 abstract 1261
 "Ähnlichkeit zwischen Vokalforman-
 ten und Formanten von Musikin-
 strumenten" 1261
 "Dynamic Spectrograms of Speech"
 771
 "Selective Voice Control Problems"
 abstract 771
 "Similarity Between Vowel Formants
 and the Formants of Musical Instru-
 ments" 1261
- Kohman, G. T.
 "Polyethylene Terephthalate — Its
 Use as a Capacitor Dielectric"
 abstract 1264-5
- Kolb, E. D.
 "Impurity Effects in the Thermal Con-
 version of Germanium"
 abstract 257
 "Solute Distribution in Germanium
 Crystals" 1512
- Koonce, Miss S. E.
 "Growth of Metal Whiskers" 1261
- Kruger, M. K.
 biographical material 1017-18
 "L3 Coaxial System — Quality Con-
 trol Requirements" 943-67
- KRYPTON
 ions, drift velocity 517
- k-STATISTICS
 tensor notation and sampling cumu-
 lants 770
- L
- Labrie, L. J.
 "High Frequency Crystal Units for
 Use in Selective Networks and Their
 Proposed Application in Filters
 Suitable for Mobile Radio Channel
 Selection" 1508
- LAMINATED CONDUCTORS *See* Conduc-
 tor
- Lander, J. J.
 "Solubility and Diffusion Coefficient of
 Carbon in Nickel: Reaction Rates
 of Nickel-Carbon Alloys with
 Barium Oxide"
 abstract 771
- "Vacuum Tube Electrometer Ampli-
 fier" 1261
 "Large Current Amplifications in Fila-
 mentary Transistors" (W. Van Roos-
 broeck) 774
 "LARMOR FREQUENCY" 85
 "LARMOR PRECESSION" 85
 LATEX EBONITE 513
 "Lead-Acid Stationary Batteries" (U. B.
 Thomas)
 abstract 257.
- Lebert, A. W.
 "Standardization of Rigid Coaxial
 Transmission Lines" 1009
- LENTINUS LEPIDEUS
 Mad. 534 (test fungus) 154
- LENZITES TRABEA
 Mad. 617 (test fungus) 155
- Lewis, H. W.
 "Calibration of the Rolling Ball Vis-
 cometer" 1010
 "Multiple Meson Production in Nuc-
 leon-Nucleon Collisions" 771
 "Spherical Model of a Ferromagnet"
 abstract 510
 "Lightning Protection for Mobile Radio
 Fixed Stations" (D. W. Bodle)
 abstract 257
- LINE INSULATION *See* Automatic Line
 Insulation Test
- Linville, J. G.
 "Transistor Negative-Impedance Con-
 verters" 1510
- LIQUID(s)
 magnetic resonance 384
- LITHIUM
 diffusion into germanium and silicon
 1509
- Livadary, J. P.
 "Dual Photomagnetic Intermediate
 Studio Recording"
 abstract 769
- Logan, Mason A.
 biographical material 1515
 "Dynamic Measurements on Electro-
 magnetic Devices" 1413-67
- Lovell, G. H.
 biographical material 1018
 "L3 Coaxial System — Amplifiers"
 879-914

- "Low-Drain Transistor Audio Oscillator" (D. E. Thomas)
abstract 516
- "Low-Loss Waveguide Transmission" (Miller and Beck)
abstract 1011
- L3 COAXIAL SYSTEM**
- Amplifiers (Morris, Lovell and Dickin-
son) 879-914
- Application of Quality Control Re-
quirements in the Manufacture of
Components (Garrett, Tuffnell and
Waddell) 969-1005
- Equalization and Regulation (Ketch-
ledge and Finch) 833-78
- Foreword (E. I. Green) 779-80
- Quality Control Requirements (Dodge,
Kinsburg and Kruger) 943-67
- System Design (Elmendorf, Ehrbar,
Klie and Grossman) 781-832
- Television Terminals (Rieke and Gra-
ham) 915-42
- LUBRICANT(S)**
solid 1508
- Luke, C. L.
"Photometric Determination of Anti-
mony in Lead Using the Rhodamine
B Method" 1261
- "Photometric Determination of Silicon
in Ferrous, Ferromagnetic, Nickel,
and Copper Alloys — A Molyb-
denum Blue Method"
abstract 772
- Lumsden, G. Q.
"Quarter Century of Evaluating Pole
Preservatives" 772
- "Review of American Standard Fiber
Stresses of Wood Poles" 1262
- M**
- McKay, K. G.
"Bombardment Conductivity" 1510
- "Crystal Conduction Counter"
abstract 1262
- McLean, D. A.
"Microstructure of Capacitor Paper"
1510
- McMahon, W.
"Polyethylene Terephthalate — Its
Use as a Capacitor Dielectric"
abstract 1264-5
- McMillan, B.
"Basic Theorems of Information The-
ory"
abstract 1510
- McRae, J. W.
biographical material 776
- "Solderless Wrapped Connections —
Introduction" 523-4
- "Transistors in Our Civilian Econ-
omy"
abstract 510
- MacWilliams, W. H., Jr.
"Computers — Past, Present, and Fu-
ture"
abstract 772
- MAGNETIC ALLOYS** 1155, 1258, 1262
- "Magnetic Crystal Anisotropy and Mag-
netostriction of Iron-Nickel Alloys"
(Bozorth and Walker)
abstract 1008-9
- "Magnetic Double Refraction at Micro-
wave Frequencies" (Weiss and Fox)
517
- MAGNETIC FIELD**
noise in a germanium filament 647
- MAGNETIC MATERIALS**
behavior 1259
- classical concepts 1155
- phenomena, new 1155-72
- MAGNETIC RESONANCE**
discovery and importance 74
- equation, general 80
- nuclear 74-99
- practical value 74
- "Magnetic Resonance of Electrons"
(Karl K. Darrow) 384-405
- MAGNETIC STRUCTURE**
Mishima alloys 1262
- "Magnetic Study of Low Temperature
Transformation in Magnetite" (Wil-
liams and Bozorth) 1264
- MAGNETITE**
low temperature transformation, mag-
netic study 1264
- "Magnetomechanical Effects in an Anti-
ferromagnet CoO" (M. E. Fine)
1260

- "Magnetoresistance Effect in InSb"
(Pearson and Tanenbaum) 1263
- Mallina, R. F.
biographical material 775-6
- "Solderless Wrapped Connections —
Part I — Structure and Tools"
525-55
- Mallinckrodt, C. O.
"Experimental Verification of the The-
ory of Laminated Conductors"
abstract 255
- Mapes, C. M.
"Telephone System in National De-
fense"
abstract 258
- Mason, W. P.
biographical material 776
- "Hall Effect Modulators and 'Gyrat-
ors' Employing Magnetic Field
Independent Orientations in Ger-
manium"
abstract 1010
- "Piezoelectric, Dielectric, and Elastic
Properties of $\text{ND}_4\text{D}_2\text{PO}_4$ "
abstract 510
- "Rotational Relaxation in Nickel at
High Frequencies"
abstract 1262
- "Solderless Wrapped Connections —
Part II — Necessary Conditions for
Obtaining a Permanent Connection"
557-90
- "Matching Coaxial Line to the Ground-
Plane Antenna" (R. T. DeCamp)
abstract 258
- MATHIEU'S EQUATION
connection formulas between the solu-
tions of 1014
- Mathy, P. T.
"Conveyorization Avoids Handling
Headaches" 1262
- Matthias, B. T.
"Piezoelectric, Dielectric, and Elastic
Properties of $\text{ND}_4\text{D}_2\text{PO}_4$ "
abstract 510
- "Polymorphism of $\text{ND}_4\text{D}_2\text{PO}_4$ "
abstract 257
- "Superconducting Properties of Cobalt
Disilicide"
abstract 1011
- "Superconductivity in the Cobalt-Sili-
con System"
abstract 256
- May, A. S.
"Microwave System Test Equipment"
1510
- "Mean Free Paths of Electrons in Evap-
orated Metal Films" (Reynolds and
Stilwell) 514
- "Measurement of Magnetostriction in
Single Crystals" (Bozorth and Ham-
ming)
abstract 1008
- "Measurement of Minority Carrier Life-
time in Germanium" (L. B. Valdes)
abstract 517
- MECHANICAL STABILITY
connections, solderless, wrapped 591
- "Mechanized Billing of AMA Toll Mes-
sages" (F. D. Slade) 1264
- MELLOR APPROXIMATION 100, 111-13
- Mertz, P.
"Influence of Echoes on Television
Transmission" 1511
- Merz, W. J.
"Domain Properties in BaTiO_3 "
abstract 511
- "Polymorphism of $\text{ND}_4\text{D}_2\text{PO}_4$ "
abstract 257
- MESON PRODUCTION
in nucleon-nucleon collisions 771
- MESSAGE ACCOUNTING *See* Automatic
Message Accounting
- METAL
whiskers, growth 1261
zone-melting 513
- METALLURGY
solid state physics 257
- "Methods in Electron Microscopy of
Solids" (R. D. Heidenreich)
abstract 770
- MICROSCOPY
electron, of solids, methods 770
- "Microstructure of Capacitor Paper"
(McLean, Birdsall and Calbick)
1510

MICROWAVE FREQUENCIES

- ferrites 1333-69
- ferromagnetic Faraday effect 1260
- magnetic double refraction 517

MICROWAVE NOISE

- in electron streams, calculation 773
- "Microwave Observation of the Collision Frequency of Electrons in Germanium" (Benedict and Shockley) 1258

MICROWAVE RADIO

- routes, path testing 1508

MICROWAVE RELAY STATIONS

- antenna 770
- "Microwave System Test Equipment" (A. S. May) 1510
- "Microwaves from Coast-to-Coast" (J. R. Rae) 1013

Miller, E. S.

- "Notes on Methods of Transmitting the Circular Electric Wave Around Bends"
- abstract 511

Miller, R. L.

- "Auditory Tests with Synthetic Vowels"
- abstract 1011
- "Dynamic Spectrograms of Speech" 771

Miller, S. E.

- "Low-Loss Waveguide Transmission"
- abstract 1011
- "Multi-Element Directional Couplers"
- abstract 258

MISHIMA ALLOYS 1262

MOBILE RADIO *See* Radio

- "Mobility in High Electric Fields" (E. M. Conwell)
- abstract 768
- "Mobility of Electrons in Germanium" (Debye and Conwell)
- abstract 507

- "Mobility of Holes and Electrons in High Electric Fields" (E. J. Ryder)
- abstract 1511-12

MODULATOR(S)

- employing magnetic field independent orientations in germanium 1010

MOISTURE

- insulation of subscriber lines 627

MOLYBDENUM BLUE METHOD 772

Montgomery, H. C.

- "Transistor Noise in Circuit Applications"
- abstract 511-12

Morgan, S. P.

- "Experimental Verification of the Theory of Laminated Conductors"
- abstract 255

Morris, Lester H.

- biographical material 1018
- "L3 Coaxial System — Amplifiers" 879-914

Morrison, J.

- "Controlled Gas Leak" 1262
- "Motion of Domain Walls in Ferrite Crystals" (Galt, Andrus and Hopper) 1260
- "Motion of Gaseous Ions in Strong Electronic Fields" (Gregory H. Wannier) 170-254

- abstract 259

MOTION PICTURES

- three-phase power from single-phase source 256

"Multi-Element Directional Couplers" (Miller and Mumford)

- abstract 258, 511

"Multiple Meson Production in Nucleon-Nucleon Collisions" (H. W. Lewis) 771

Mumford, W. W.

- "Multi-Element Directional Couplers"
- abstract 258, 511

"Optimum Piston Position for Wide-Band Coaxial-to-Waveguide Transducers"

- abstract 772

Munson, W. A.

- "In Search of the Missing 6 Db"
- abstract 512

Murray, G. E.

- biographical material 1269
- "Development and Manufacture of Electro-formed Conductor for Telephone Drop Wire" 1099-1135

N

- NATIONAL DEFENSE *See* Civil defense
- NATIONAL POLITICAL CONVENTIONS (Chicago)
television coverage 1263
- "Nation-Wide Numbering Plan" (W. H. Nunn)
abstract 512
- "Nature of Solids" (G. H. Wannier)
abstract 517
- $\text{ND}_4\text{D}_2\text{PO}_4$ *See* Ammonium dihydrogen phosphate
- Nelson, R. A.
"Vapor Pressure of He^3 α He^4 Mixtures" 772
- Nesbitt, E. A.
"Physical and Magnetic Structure of of the Mishima Alloys" 1262
- "New General-Purpose Relay for Telephone Switching Systems" (A. C. Keller)
abstract 510
- "New Infrared Absorption Bands in p -Type Germanium" (Briggs and Fletcher)
abstract 507
- "New Method of Calculating Microwave Noise in Electron Streams" (J. R. Pierce)
abstract 773
- "New Non-Reciprocal Waveguide Medium Using Ferrites" (E. H. Turner) 1512
- Newsom, James B.
biographical material 1269-70
"Card Translator for Nationwide Dialing" 1037-98
- NICKEL
carbon in, solubility and diffusion coefficient 771
rotational relaxation at high frequencies 1262
- NICKEL ALLOYS
silicon, photometric determination 772
- NICKEL-CARBON ALLOYS
barium oxide, reaction rates with 771
- NITROGEN
ions, drift velocity 1013

NOISE

- germanium filament, theory of magnetic effect on 647-64
- in a temperature-limited electron beam in a drift space 773
- microwave, in electron streams, calculation 773
- oscillatory 1371
- "Non-Blocking Switching Networks Study" (Charles Clos) 406-24
- "Nonsynchronous Time Division with Holding and with Random Sampling" (Pierce and Hopper)
abstract 258, 513
- "Notes on Methods of Transmitting the Circular Electric Wave Around Bends" (E. S. Miller)
abstract 511
- N-P-N JUNCTION PHOTOTRANSISTORS
See Phototransistor
- NUCLEAR MAGNETIC RESONANCE
discovery 99
- "Nuclear Magnetic Resonance" (Karl K. Darrow) 74-99
- NUCLEON-NUCLEON COLLISIONS
multiple Meson production 771
- NUCLEUS, nuclei
magnetic moments 74
- NUMBER(s), numbering *See* Telephone numbers
- Nunn, W. H.
"Nation-Wide Numbering Plan"
abstract 512
- Nye, J. F.
"Geometrical Relations in Dislocated Crystals"
abstract 1012

O

- O'Connor, S. F.
"Plating Room Waste Water Disposal" 1012
- OIL-TYPE PRESERVATIVES 160
- Olmstead, P. S.
"How to Detect the Type of an Assignable Cause"
abstract 512
- Olsen, K. M.
"Purification and Prevention of Segre-

- gation in Single Crystals of Germanium" 1012
- "On the Band Width of Vowel Formants" (B. P. Bogert) abstract 1507-8
- "On the Number of Symmetry Types of Boolean Functions on n Variables" (D. Slepian) 1264
- "Optical Position Encoder and Digit Register" (Follingstad, Shive and Yaeger) abstract 508-9
- "Optimum Piston Position for Wide-Band Coaxial-to-Waveguide Transducers" (W. W. Mumford) abstract 772
- "Organization of the Engineering Profession" (D. A. Quarles) abstract 514
- Osborne, H. S.
 "Automatic Switching for National Wide Telephone Service" abstract 507
 "Rose by Any Other Name" 773
- OSCILLATOR
 low-drain transistor audio 516
 transistor, use in multifrequency pulsing current supply 1313-31
- Osmer, T. F.
 biographical material 776
 "Solderless Wrapped Connections — Part II — Necessary Conditions for Obtaining a Permanent Connection" 557-90
- OVERSEAS TELEPHONY
 single-sideband system 514
- Owens, C. D.
 "Analysis of Measurements on Magnetic Ferrites" abstract 1012
- OXYGEN
 ions, drift velocity 1013
- P**
- PARAMAGNETIC RESONANCE
 exchange narrowing 1257
- PARAMAGNETIC SALTS
 magnetic resonance 74
- PARAMAGNETIC SOLIDS
 electron resonance 393
- "Path Testing for Microwave Radio Routes" (R. D. Campbell) 1508
- Pearson, G. L.
 "Magnetoresistance Effect in InSb" 1263
 "Silicon p-n Junction Alloy Diodes" abstract 512-13
- Peck, D. S.
 "Ten-Stage Cold-Cathode Stepping Tube" abstract 773
- PENTACHLOROPHENOL
 evaluation 488
- "Permalloy Problem" (R. M. Bozorth) abstract 1258-9
- Peters, H.
 "Hard Rubber" abstract 513
- Peterson, G. E.
 "Application of Information Theory to Research in Experimental Phonetics" abstract 513
- "Basic Physical Systems for Communication Between Two Individuals" 1511
- "Information-Bearing Elements of Speech" abstract 773
- Pfann, W. G.
 "Principles of Zone-Melting" abstract 513
- "Purification and Prevention of Segregation in Single Crystals of Germanium" 1012
- "Segregation of Two Solutes, with Particular Reference to Semiconductors" abstract 258
- PHONETICS
 information theory application 513
- PHOTOCELL(s)
 germanium, single-crystal 257
 M-1740 p-n junction 514
- PHOTOGRAPHY
 high-speed, economics 771

- PHOTOMAGNETIC INTERMEDIATE STUDIO
RECORDING 769
- "Photometric Determination of Antimony in Lead Using the Rhodamine B Method" (C. L. Luke) 1261
- "Photometric Determination of Silicon in Ferrous, Ferromagnetic, Nickel, and Copper Alloys — A Molybdenum Blue Method" (C. L. Luke) abstract 772
- PHOTOTRANSISTOR(s)
- germanium 1263
- n-p-n junction multiplier 1263
- p-n junction 1263
- point contact 1263
- See also Transistor
- "Physical and Magnetic Structure of the Mishima Alloys" (Nesbitt and Heidenreich) 1262
- Pierce, J. R.
- "New Method of Calculating Microwave Noise in Electron Streams" abstract 773
- "Nonsynchronous Time Division with Holding and with Random Sampling" abstract 258, 513
- "Transistors" 1511
- "Piezoelectric, Dielectric, and Elastic Properties of $\text{ND}_4\text{D}_2\text{PO}_4$ " (Mason and Matthias) abstract 510
- Pike, V. B.
- "Improved Electrolysis Switch" abstract 259
- Pilliod, J. J.
- "Fundamental Plans for Toll Telephone Plant" 513
- PINE (wood) See Southern pine
- PISTON
- transducers, wide-band coaxial-to-waveguide, optimum position 772
- PLANE WAVE THEORY 1333
- "Plans for Toll Telephone Plant, Fundamental" (J. J. Pilliod) 513
- PLASTICS
- stress relation 517-18
- "Plating Room Waste Water Disposal" (S. F. O'Connor) 1012
- P-N JUNCTION PHOTOTRANSISTORS See Phototransistor
- P-N JUNCTION SILICON DIODE 512
- P-N-P-N TRANSISTORS See Transistor
- POINT CONTACT PHOTOTRANSISTORS See Phototransistor
- POLE(s) See Telephone poles
- Pollard, C. E., Jr.
- "Balanced Polar Mercury Contact Relay" 1393-1411
- biographical material 1516
- POLYETHYLENE
- connections, solderless, wrapped 557
- "Polyethylene Insulated Telephone Cable" (A. S. Windeler) 1245-56
- "Polyethylene Terephthalate — Its Use as a Capacitor Dielectric" (Wooley, Kohman and McMahon) abstract 1264-5
- "Polymorphism of $\text{ND}_4\text{D}_2\text{PO}_4$ " (Wood, Merz and Matthias) abstract 257
- Portis, A. M.
- "Electron Spin Resonance in a Silicon Semi-Conductor" 1511
- "Practice of Quality Control" (A. G. Dalton) abstract 1009
- PRECESSION 84
- Prim, R. C.
- biographical material 776-7
- "DC Field Distribution in a 'Swept Intrinsic' Semiconductor Configuration" 665-94
- "Space-Charge Limited Emission in Semi-Conductors" abstract 1512
- Prince, M. B.
- "Interpretation of α -values in p-n Junction Transistors" abstract 1009
- "Principles and Applications of Converters for High-Frequency Measurements" (D. A. Alsberg) abstract 506
- "Principles of Tape-to-Card Conversion in the AMA System" (W. B. Groth) 1260

"Principles of Zone-Melting" (W. G. Pfann)

abstract 513

"Progress and Problems" (D. A. Quarles) 1511

"Properties of Germanium Phototransistors" (J. N. Shive)

abstract 1263

"Properties of Junction Transistors (K. D. Smith) 774

"Properties of M-1740 p-n Junction Photocells" (J. N. Shive)

abstract 514

"Properties of Silicon and Germanium" (Esther M. Conwell)

abstract 507

"Properties of Thermally Produced Acceptors in Germanium" (Fuller, Theuerer and Van Roosbroeck)

abstract 256

PROTON RESONANCE ABSORPTION 75

Pugh, S. G.

"Southern Bell Switches to Chemical Brush Control"

abstract 1263

"Purification and Prevention of Segregation in Single Crystals of Germanium" (Pfann and Olsen) 1012

Q

QUALITY CONTROL

L3 coaxial system 943, 969

Quarles, Donald A.

"AIEE Progress"

abstract 1013

"Organization of the Engineering Profession"

abstract 514

"Progress and Problems" 1511

"Report to the Membership" 1511

"We Begin a New Institute Year"

abstract 514

"Quarter Century of Evaluating Pole Preservatives" (G. Q. Lumsden)

772

R

RADIATION

out-of-band 514

RADIO(s)

connections, solderless, wrapped 525 ff

field strength fluctuations far beyond the horizon 773

frequency spectrum, use 42

lightning protection 257

microwave, routes, path testing 1508

mobile:

bands, frequency economy 42-62, 1508

channel selection, filters suitable for 1508

channels usability 42-62

fixed stations, lightning protection 257

intermodulation interference 63-73

path loss between antennas 43

transmission, comparison at 150, 450, 900 and 3700 Mc 1513

super-high frequency, signal fading, study 1187-1202

"Radio Transmission Beyond the Horizon in the 40- to 4,000-Mc Band" (K. Bullington)

abstract 767-8

Rae, J. B.

"Microwaves from Coast-to-Coast" 1013

Ralston, R. W.

"Television Coverage of the National Political Conventions"

abstract 1263

RANDOM SERVICE

calls, delayed exponential, working curves 360-83

delay curves 100-19

RAPID LINE INSULATION TESTING

switching systems 627, 631-3

Read, W. T.

"Anisotropic Elasticity with Applications to Dislocation Theory"

abstract 1509

"Statistics of the Recombinations of Holes and Electrons"

abstract 259

RECORDING *See* Sound recording

RECTIFIER(s)

germanium, single-crystal 257

selenium 1469-92

- Reed, E. D.
biographical material 777
"Coupled Resonator Reflex Klystron"
715-66
- REFLECTOR
plane, antenna used with 770
- REGENERATIVE AMPLIFIER *See* Amplifier
- "Regenerative Amplifier for Digital Computer Applications" (J. H. Felker)
abstract 508
- RELAXATION
electric fields, high, mobility in 768
ferromagnetic materials 767
nickel at high frequencies, rotational 1262
nuclear magnetic resonance 92-9
- RELAY
AF type 510
AG 510
AJ 510
electromagnetic, general purpose 510
mercury contact, balanced polar 1393-1411
"Report to the Membership" (D. A. Quarles) 1511
- RESONANCE 83
dimensional 1163-4
electron spin, in a silicon semi-conductor 1511
ferromagnetic 1164-6
See also Magnetic resonance; Nuclear magnetic resonance; Paramagnetic resonance
- RESONATOR
coupled reflex klystron 715-66
"Review of American Standard Fiber Stresses of Wood Poles" (Lumsden and Colley) 1262
"Review of New Magnetic Phenomena" (R. E. Alley, Jr.) 1155-72
- Reynolds, F. W.
"Mean Free Paths of Electrons in Evaporated Metal Films" 514
- RHODAMINE B METHOD
photometric determination of anti-mony in lead 1261
- Rice, S. O.
"Statistical Fluctuations of Radio Field Strength Far Beyond the Horizon"
abstract 773
- Rieke, John W.
biographical material 1018
"L3 Coaxial System — Television Terminals" 915-42
- Riordan, John
biographical material 263
"Delay Curves for Calls Served at Random" 100-19
addendum 1266
- Ritchie, Alistair E.
biographical material 520-1
"Throwdown Machine for Telephone Traffic Studies" 292-359
"Rose by Any Other Name" (H. S. Osborne) 773
"Rotational Relaxation in Nickel at High Frequencies" (W. P. Mason)
abstract 1262
- Rowen, John H.
biographical material 1516
"Ferrites in Microwave Applications" 1333-69
- RUBBER(s)
conductive 1261
hard 513
synthetic 513
- Rudisill, J. A., Jr.
"Simple Phase-Angle Measurement Technique"
abstract 257
- Ryder, E. J.
"Mobility of Holes and Electrons in High Electric Fields"
abstract 1511-12
- S
- SALT (paramagnetic)
magnetic resonance 74
- Sawyer, B.
"Silicon p-n Junction Alloy Diodes"
abstract 512-13
- Schimpf, L. G.
"High-frequency Transistor Tetrode"
abstract 774

- "Junction Transistor Tetrode for High-Frequency Use"
abstract 517
- Schlaak, N. F.
"Single-Sideband System for Overseas Telephony"
abstract 514
- "Science and Scientists in Telecommunications" (R. M. Burns) 1259
- "Segregation of Two Solutes, with Particular Reference to Semiconductors" (W. G. Pfann)
abstract 258
- "Selective Voice Control Problems" (W. E. Kock)
abstract 771
- "Selenium Rectifiers — Factors in Their Application" (J. Gramels) 1469-92
- SEMICONDUCTOR(s)
DC field in a "swept intrinsic" 665-94
diode gates 1137-54
silicon, electron spin resonance 1511
solute segregation 258
space-charge limited emission 1512
statistics of the recombination of holes and electrons 259
"swept intrinsic" configuration, DC field distribution in 665-94
- "Semiconductor Diode Gates" (Luther W. Hussey) 1137-54
- SERIAL ADDER
transistor 509
- SHIFT REGISTER
transistor 509
- Shipley, F. F.
"Automatic Toll Switching Systems" 514
- Shive, J. N.
"Optical Position Encoder and Digit Register"
abstract 508-9
- "Properties of Germanium Phototransistors"
abstract 1263
- "Properties of M-1740 p-n Junction Photocells"
abstract 514
- Shockley, William
"Anisotropic Elasticity with Applications to Dislocation Theory"
abstract 1509
- "Interpretation of e/m Values for Electrons in Crystals" 515
- "Microwave Observation of the Collision Frequency of Electrons in Germanium" 1258
- "Solid State Physics in Electronics and in Metallurgy"
abstract 257
- "Space-Charge Limited Emission in Semi-Conductors"
abstract 1512
- "Statistics of the Recombinations of Holes and Electrons"
abstract 259
- "Transistor Electronics: Imperfections, Unipolar and Analog Transistors"
abstract 515
- "Unipolar 'Field-Effect' Transistor"
abstract 515
- SHORT ARC
characteristics 1493-1506
- SILICON
carrier density and mobility 507
electrical properties 507
electrons and holes, field dependence of mobility 1511
lithium diffusion 1509
p-n junction alloy diodes 512-13
properties 507
semi-conductor, electron spin resonance 1511
temporary traps 1260
varistors 514
- "Silicon p-n Junction Alloy Diodes" (Pearson and Sawyer)
abstract 512-13
- "Similarity Between Vowel Formants and the Formants of Musical Instruments" (W. E. Kock) 1261
- "Simple Attachment for Low Temperature Use of an X-Ray Diffraction Camera" (E. A. Wood) 1264
- "Simple Phase-Angle Measurement Technique" (J. A. Rudisill, Jr.)
abstract 257

- "Single-Crystal Germanium" (Teal, Sparks and Buehler)
abstract 257
- SINGLE-PHASE 115-VOLT POWER
conversion to three-phase 230-volt form 256
- "Single-Sideband System for Overseas Telephony" (N. F. Schlaak)
abstract 514
- Slade, F. D.
"Mechanized Billing of AMA Toll Messages" 1264
- Slepian, D.
"On the Number of Symmetry Types of Boolean Functions on n Variables" 1264
- Slichter, W. P.
"Impurity Effects in the Thermal Conversion of Germanium"
abstract 257
"Solute Distribution in Germanium Crystals" 1512
- Smith, K. D.
"Properties of Junction Transistors" 774
- SOIL-BLOCK TESTS
creosotes, residual, characteristics, of weathered blocks 479-87
evaluation 132-60
- SOLDER JOINT 525
- SOLDERLESS WRAPPED CONNECTIONS 523-610
"Solderless Wrapped Connections"
Introduction (J. W. McRae) 523-4
Part I—Structure and Tools (R. F. Mallina) 525-55
Part II—Necessary Conditions for Obtaining a Permanent Connection (Mason and Osmer) 557-90
Part III—Evaluation and Performance Tests (R. H. Van Horn) 591-610
- SOLID(s)
magnetic resonance 384
nature of 517
paramagnetic, electron resonance 393
"Solid Lubricants" (W. E. Campbell) 1508
- "Solid State Physics in Electronics and in Metallurgy" (W. Shockley)
abstract 257
- "Solubility and Diffusion Coefficient of Carbon in Nickel: Reaction Rates of Nickel-Carbon Alloys with Barium Oxide" (Lander, Kern and Beach)
abstract 771
- SOLUTE(s)
germanium crystals, distribution 1512
segregation during directional solidification of an ingot 258
"Solute Distribution in Germanium Crystals" (Slichter and Kolb) 1512
- SOUND PRESSURE
ear canal 512
- SOUND RECORDING
dial photomagnetic intermediate studio recording 769
magnetic 509
mechanical 509
photographic 509
processes comparison 509
re-recording 769
three-phase power from single-phase source 256
- "Southern Bell Switches to Chemical Brush Control" (S. G. Pugh)
abstract 1263
- SOUTHERN PINE SAPWOOD
stakes, testing 427
- "Space-Charge Limited Emission in Semi-Conductors" (Shockley and Prim)
abstract 1512
- "Space-Charge Limited Hole Current in Germanium" (G. C. Dacey)
abstract 1509
- Sparks, M.
"Junction Transistor"
abstract 516
"Single-Crystal Germanium"
abstract 257
- SPECTROGRAM(s)
speech 771
- SPEECH
information-bearing elements 773
phonetic content 771
spectrograms 771

"Spherical Model of a Ferromagnet"
(Lewis and Wannier)
abstract 510

SPIN-LATTICE RELAXATION *See* Relaxation

"Standardization of Rigid Coaxial Transmission Lines" (Drvestep and Lebert) 1009

Stang, L. R.

"Telephone Answering Services" 516

Stansel, F. R.

"Transistor Equations"
abstract 1013

"Statistical Fluctuations of Radio Field Strength Far Beyond the Horizon"
(S. O. Rice)
abstract 773

"Statistical Study of Selective Fading of Super-High Frequency Radio Signals" (R. L. Kaylor) 1187-1202

"Statistics of the Recombinations of Holes and Electrons" (Shockley and Read)
abstract 259

STERILIZATION

wood preservation 152

Stewart, J. A.

"Traffic Engineering Design of Dial Telephone Exchanges" 516
abstract 257

Stilwell, G. R.

"Mean Free Paths of Electrons in Evaporated Metal Films" 514

Stone, H. A., Jr.

biographical material 521

"Ferrite Core Inductors" 265-91

STRAIN(s)

solderless wrapped connections 557

STRESS(es)

solderless wrapped connections 557

"Stress Relaxation in Plastics and Insulating Materials" (E. E. Wright)
abstract 517-18

Struthers, J. D.

"Copper as an Acceptor Element in Germanium"
abstract 256

Stubner, F. W.

"Acceleration Effects on Electron Tubes" 1203-29

biographical material 1270

Suhl, Harry

biographical material 777

"Theory of Magnetic Effects on the Noise in a Germanium Filament" 647-64

SULFUR HEXAFLUORIDE

ions, negative, formative of 1007

"Superconducting Properties of Cobalt Disilicide" (Matthias and Hulm)
abstract 1011

"Superconductivity in the Cobalt-Silicon System" (B. T. Matthias)
abstract 256

"Surface Properties of Germanium" (Brattain and Bardeen) 1-41

SWEDISH CREOSOTE EVALUATION TESTS 489-90

SWITCHES

dynamic measurements of reciprocating phenomena 1413
electrolysis, improved 259

SWITCHING CIRCUITS

arcing of electrical contacts 1231-44, 1493-1506
contact erosion 1493

SWITCHING SYSTEMS

AF-type relay 510

automatic 507, 514

balanced polar mercury contact relay 1393

calls delay 100

cold cathode gas filled tubes 1371

counting, reversible binary 516

crossbar No. 5, traffic study 292-359

crossnet array 406

crosspoints 406, 419

ferroelectric storage elements 506

nation-wide telephone service 507, 1019-36

non-blocking networks 406-24

rapid line insulation testing 627, 631-3

toll 514, 1019-36

transistors 506

voice frequency signals 1371

T

Tanenbaum, M.

"Magnetoresistance Effect in InSb"
1263

TAPE-TO-CARD CONVERSION
in the AMA system 1260

Teal, G. K.

"Single-Crystal Germanium"
abstract 257

Tebo, Julian D.

appointed editor *opp* 523

TELECOMMUNICATION(s)

science and scientists in 1259

"Telephone Answering Services" (L. R.
Stang) 516

TELEPHONE CALLS *See* Call(s)

"Telephone Industry in National De-
fense" (C. A. Armstrong)
abstract 507

TELEPHONE NUMBERS

nation-wide plan 512

TELEPHONE POLES, Posts

fiber stresses 1262

preservatives, evaluation 772

testing 425-505

TELEPHONE SET (500 type)

circuit, improved 611-26

"Telephone System in National De-
fense" (C. M. Mapes)

abstract 258

TELEPHONE TRAFFIC

delay formula 100

throwdown machine 292-359

"Telephony Development in the United
States" (A. B. Clark)

abstract 768

TELEVISION

connections, solderless, wrapped
525 ff

political conventions, Chicago 1263

terminals, L3 coaxial system 915-42

transmission, influence of echoes 1511

"Television Coverage of the National
Political Conventions" (Ralston and
Wickline)

abstract 1263

"Temporary Traps in Silicon and Ger-
manium" (Haynes and Hornbeck)
1260

"Tensor Notation and the Sampling
Cumulants of k -Statistics" (E. L.
Kaplan)

abstract 770

"Ten-Stage Cold-Cathode Stepping
Tube" (D. S. Peck)

abstract 773

TERMINAL(s)

connections, solderless, wrapped 523-
610

TETRAGONAL CRYSTAL $\text{ND}_4\text{D}_2\text{PO}_4$

properties 510

TETRODE(s)

transistor, high-frequency 774

transistor, junction, for high-fre-
quency use 517

"Theoretical Study of an Antenna-Re-
flector Problem" (W. C. Jakes, Jr.)
abstract 770

"Theory of Magnetic Effects on the
Noise in a Germanium Filament"
(Harry Suhl) 647-64

Theuerer, H. C.

"Properties of Thermally Produced
Acceptors in Germanium"

abstract 256

Thomas, D. E.

"Low-Drain Transistor Audio Oscil-
lator"

abstract 516

"Transistor Amplifier - Cutoff Fre-
quency"

abstract 516

Thomas, U. B.

"Lead-Acid Stationary Batteries"

abstract 257

"Three-Phase Power From Single-Phase
Source" (A. L. Holcomb)

abstract 256

"Threshold Law for Single Ionization of
Atoms or Ions by Electrons" (G. H.
Wannier)

abstract 1512-13

"Throwdown Machine for Telephone
Traffic Studies" (Frost, Keister and
Ritchie) 292-359

THROWDOWN TESTS

call delays 360

TIMBER PRESERVATION *See* Wood pres-
ervatives

- TIP AND RING GROUND TEST 636
- TOLL MESSAGES
 billing, mechanized 1264
- TOLL SWITCHING
 nationwide 1019
- TOLL TELEPHONE PLANT
 plans 513
- TOLUENE
 as a diluent for creosote treating solutions 455
- Townsend, J. R.
 "Dynamic Program for Conversion" 1512
 "What We Have Learned in 1952" 774
- Townsend, Mark A.
 biographical material 1516
 "Cold Cathode Tubes for Transmission of Audio Frequency Signals" 1371-91
- TRAFFIC *See* Telephone traffic
- "Traffic Engineering Design of Dial Telephone Exchanges" (J. A. Stewart) 516
 abstract 257
- TRANSDUCER(s)
 wide-band coaxial-to-waveguide, optimum piston position 772
- TRANSISTOR(s)
 amplifier 516
 analog 515
 audio-oscillator 516
 circuit application noise 511
 circuit cutoff frequency 516
 circuits, negative-resistance 508
 civilian economy 510
 computer, digital, block diagrams 255, 508, 769
 converters, negative-impedance 1510
 counter, reversible binary 516
 digital computer, block diagrams 255, 769
 electronics 515
 equations 1013
 field effect type 515
 filamentary, large current amplifications 774
 germanium, single-crystal 257
 imperfections 515
 junction:
 design theory 1271-1312
 principles and development 516
 properties 774
 small signal ac transmission characteristics of 1271
 space-charge layer widening 507
 tetrode for high-frequency use 517
 negative-resistance circuits 508
 noise behavior 511
 oscillator, multifrequency pulsing current supply 1313-31
 phototransistor *See* Phototransistor
 Pierce, J. R., article 1511
 p-n-p-n, four terminal 508
 point-contact:
 audio oscillator 516
 electrode spacing effect on equivalent base resistance 516
 frequency response control 515
 negative-resistance properties 508
 stability 515
 serial adder 509
 shift register 509
 tetrode, high-frequency 774
 unipolar 515
 "Transistor Amplifier — Cutoff Frequency" (D. E. Thomas)
 abstract 516
 "Transistor Electronics: Imperfections, Unipolar and Analog Transistors" (W. Shockley)
 abstract 515
 "Transistor Equations" (F. R. Stansel)
 abstract 1013
 "Transistor Negative-Impedance Converters" (J. G. Linvill) 1510
 "Transistor Noise in Circuit Applications" (H. C. Montgomery)
 abstract 511-12
 "Transistor Oscillator for Use in Multifrequency Pulsing Current Supply" (F. E. Blount) 1313-31
 "Transistor Physics Simplified" (Gentile and Barotta) 1510
 "Transistor Reversible Binary Counter" (R. L. Trent)
 abstract 516
 "Transistor Shift Register and Serial Adder" (J. R. Harris)
 abstract 509
 "Transistors" (J. R. Pierce) 1511

- "Transistors in Our Civilian Economy" (J. W. McRae)
abstract 510
- "Transistors in Switching Circuits" (A. E. Anderson)
abstract 506-7
- TRANSLATOR(s)
card, for nationwide dialing 1037
- TRANSMISSION
audio frequency signals, cold cathode tubes for 1371-91
Clogston type conductors, laminated 695-713
coaxial system L3 See Carrier: L3
electric waves, around bends 511
intertoll telephone trunks, design 1019-36
radio, beyond horizon in the 40- to 4,000-Mc band 767-8
radio, mobile, comparison at 150, 450, 900 and 3700 Mc 1513
television, influence of echoes 1511
waveguide, low-loss 1011
- "Transmission Design of Intertoll Telephone Trunks" (H. R. Huntley)
1019-36
- "Transmission Properties of Laminated Clogston Type Conductors" (E. F. Vaage) 695-713
- TRANSMITTER(s)
"Nonsynchronous Time Division with Holding and with Random Sampling" (Pierce and Hopper)
abstract 258, 513
single-sideband 514
sinusoidal radio wave 773
- "Traveling-Wave Tube" (M. E. Hines)
1510
- TRAVELING WAVE TUBE POWER CAPACITY 1509
- Trent, R. L.
"Transistor Reversible Binary Counter"
abstract 516
- TRUNK(s)
intertoll, transmission design 1019-36
- Tuffnell, T. L.
biographical material 1018
"L3 Coaxial System — Application of Quality Control Requirements in the Manufacture of Components" 969-1005
- Turner, E. H.
"New Non-Reciprocal Waveguide Medium Using Ferrites" 1512
"Type-O Carrier Telephone" (E. K. Van Tassel)
abstract 1259
"Typical Block Diagrams for a Transistor Digital Computer" (J. H. Felker)
abstract 255, 769
- U
- "Ultrasonically Induced Cavitation in Water: A Step-by-Step Process" (G. W. Willard)
abstract 1513
"Unipolar 'Field-Effect' Transistor" (W. Shockley)
abstract 515
- V
- Vaage, E. F.
biographical material 777
"Transmission Properties of Laminated Clogston Type Conductors" 695-713
- VACUUM TUBE See Electron tube
- "Vacuum Tube Electrometer Amplifier" (Hughes and Lander) 1261
- Valdes, L. B.
"Effect of Electrode Spacing on the Equivalent Base Resistance of Point-Contact Transistors"
abstract 516
"Measurement of Minority Carrier Lifetime in Germanium"
abstract 517
- Van Horn, R. H.
biographical material 777
"Solderless Wrapped Connections — Part III — Evaluation and Performance Tests" 591-610
- Van Roosbroeck, W.
"Large Current Amplifications in Filamentary Transistors" 774

- "Properties of Thermally Produced Acceptors in Germanium"
abstract 256
- Van Sielen, H. E.
"What We Did to Cut Costs of Finishing Telephone Woodwork" 1264
- Van Tassel, E. K.
"Type-O Carrier Telephone"
abstract 1259
- "Vapor Pressure of He³ and He⁴ Mixtures" (R. A. Nelson) 772
- VARISTOR(s)
copper-oxide 514
germanium 514
- Varney, R. N.
"Drift Velocities of Ions in Krypton and Xenon"
abstract 517
- "Drift Velocity of Ions in Oxygen, Nitrogen, and Carbon Monoxide"
abstract 1013
- VISCOMETER
rolling ball, calibration of 1010
- VOICE CONTROL
selective 771
- VOICE FREQUENCY SIGNALS
switching 1371
- VOWELS(s)
formants, and the formants of musical instruments 1261
formants, band width 1507-8
synthetic, auditory tests 1011

W

- Waddell, R. A.
biographical material 1018
"L3 Coaxial System — Application of Quality Control Requirements in the Manufacture of Components" 969-1005
- Walker, J. G.
"Magnetic Crystal Anisotropy and Magnetostriction of Iron-Nickel Alloys"
abstract 1008-9
- Wallace, R. L., Jr.
"High-Frequency Transistor Tetrode"
abstract 774

- "Junction Transistor Tetrode for High-Frequency Use"
abstract 517
- Wannier, Gregory H.
biographical material 263
"Connection Formulas Between the Solutions of Mathieu's Equation"
abstract 1014
"Motion of Gaseous Ions in Strong Electric Fields" 170-254
abstract 259
"Nature of Solids"
abstract 517
"Spherical Model of a Ferromagnet"
abstract 510
"Threshold Law for Single Ionization of Atoms or Ions by Electrons"
abstract 1512-13
- Warner, A. W.
"High-Frequency Crystal Units for Primary Frequency Standards"
abstract 260
- WATER
cavitation, ultrasonically induced 1513
plating room waste disposal 1012
- WAVE See Electric wave
- WAVEGUIDE(s)
directional coupler 510
multimode round, dominant wave transmission 256
non-reciprocal medium, using ferrites 1512
transmission, low-loss 1011
"We Begin a New Institute Year" (D. A. Quarles)
abstract 514
- WEATHERING
creosote and creosoted wood 457
line insulation resistance 629
wood preservatives 150
- WEIGHTED SAMPLE VALUES
approximating the mode 1010
- Weiss, M. T.
"Magnetic Double Refraction at Microwave Frequencies" 517
- Weiss, P. R.

Weiss, P. R.—*continued*:

“Exchange Narrowing in Paramagnetic Resonance”

abstract 1257-8

“Western Electric’s Service with Standards” (K. B. Clarke)

abstract 507

“What We Did to Cut Costs of Finishing Telephone Woodwork” (H. E. Van Siclen) 1264

“What We Have Learned in 1952” (J. R. Townsend) 774

Wick, R. F.

“Hall Effect Modulators and ‘Gyrators’ Employing Magnetic Field Independent Orientations in Germanium”

abstract 1010

Wickline, B. D.

“Television Coverage of the National Political Conventions”

abstract 1263

Wiener, F. M.

“In Search of the Missing 6 Db”

abstract 512

Wilkinson, Roger I.

biographical material 521

“Working Curves for Delayed Exponential Calls Served in Random Order” 360-83

Willard, G. W.

“Ultrasonically Induced Cavitation in Water: A Step-by-Step Process”

abstract 1513

Williams, H. J.

“Magnetic Study of Low Temperature Transformation in Magnetite” 1264

Windeler, A. S.

biographical material 777, 1270

“Polyethylene Insulated Telephone Cable” 1245-56

WIRE(s)

concealment 770

connections, solderless, wrapped 523-610

copper-steel, electroformed, development 1260

drop, electroformed conductor for 1099
polyethylene insulation 1245

Wood, E. A.

“Polymorphism of $\text{ND}_4\text{D}_2\text{PO}_4$ ”
abstract 257

“Simple Attachment for Low Temperature Use of an X-Ray Diffraction Camera” 1264

WOOD POLES

fiber stresses 1262

WOOD PRESERVATIVES

evaluation 120-69

WOODWORK FINISHING

cost cutting 1264

Wooley, M. C.

“Polyethylene Terephthalate — Its Use as a Capacitor Dielectric”

abstract 1264-5

“Working Curves for Delayed Exponential Calls Served in Random Order” (Roger I. Wilkinson) 360-83

WRAPPING

connections 543-5, 551

Wright, E. E.

“Stress Relaxation in Plastics and Insulating Materials”

abstract 517-18

Wright, S. B.

“Higher Frequencies for Ground-Air Communications” 1265

X

X-RAY DIFFRACTION CAMERA

attachment for low temperature use
1264

Xenon

ions, drift velocity 517

Y

Yaeger, R. E.

“Optical Position Encoder and Digit Register”

abstract 508-9

Young, W. R.

“Comparison of Mobile Radio Transmission at 150, 450, 900 and 3700 Mc”
1513

Z

ZONE-MELTING PRINCIPLES 513

Printed in U.S.A.

